Dataetisk metode & teori

Dataetisk metode & teori

Oktober 2021

Analyse af Frej Klem Thomsen

& Lizette Ingemann Skou Olsen

Publikationen er udarbejdet af

Institut for Menneskerettigheder

til Dataetisk Råd

Indholdsfortegnelse

Kapitel 1

Hvad er dataetik?

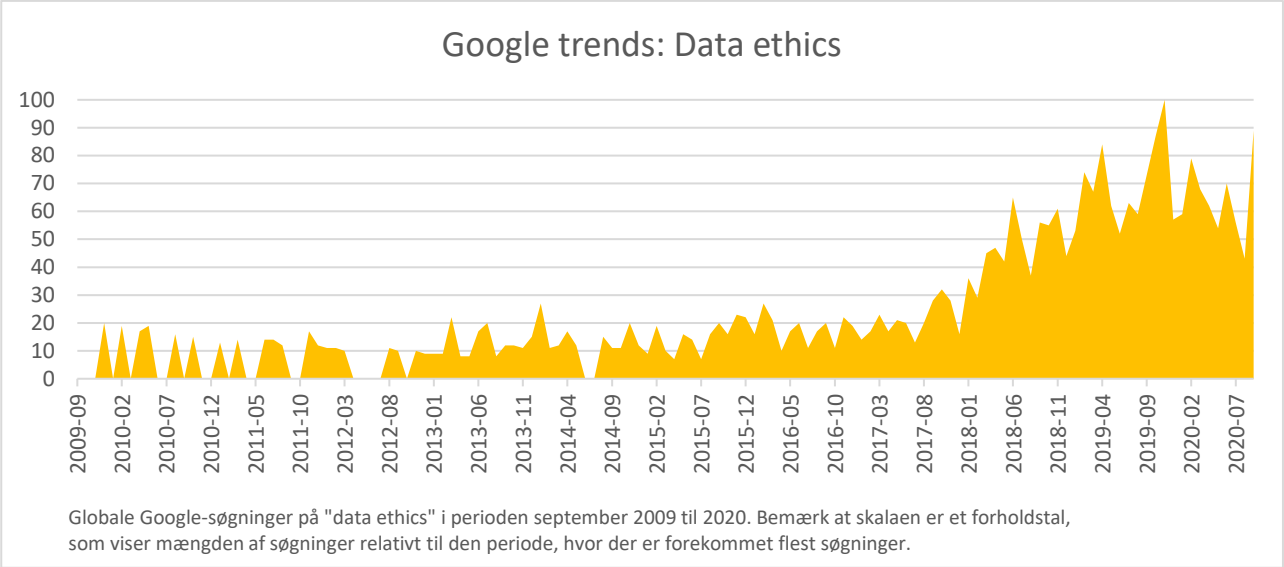
Hvad er dataetik?

Dataetisk Råd arbejder med at analysere og skabe debat om dataetik. I disse år oplever Danmark en hastig vækst i indsamling og anvendelse af data, både i den private og den offentlige sektor. For at kunne forstå og styre udviklingen har vi brug for at reflektere over dataetikken. Analyser af og debat om dataetik er således afgørende for, at Danmark kan være på forkant med den teknologiske udvikling. Men hvad er dataetik?

Dataetik er et relativt nyt begreb, som først de seneste år har tiltrukket sig bred opmærksomhed (se figur nedenfor). Af samme grund er dataetik ikke veldefineret: der findes ikke en klar og ukontroversiel etableret betydning. Dataetisk Råd har fået udarbejdet denne analyse for at afklare og præcisere, hvordan dataetik kan forstås.

> *"Dataetik forstås overordnet som den etiske dimension af forholdet mellem på den ene side teknologi og på den anden side borgernes grundlæggende rettigheder, retssikkerhed og grundlæggende samfundsmæssige værdier, som den teknologiske udvikling giver anledning til at overveje."*
>
> *Dataetisk Råds kommissorium*

Analysen anvender begrebsanalyse på baggrund af et litteraturstudie. Begrebsanalyse anvendes i forskningen i etik og politisk teori som metode til at formulere klare og præcise definitioner af relevante begreber. Analysen skitserer derfor indledningsvis, hvad en definition er og to forskellige måder at definere et begreb på.



Globale Google-søgninger på "data ethics" i perioden september 2009 til 2020. Bemærk at skalaen er et forholdstal, som viser mængden af søgninger relativt til den periode, hvor der er forekommet flest søgninger.

Analysen præsenterer der på et overblik over, hvordan dataetik defineres i forsknings- og policy-sammenhænge. Denne del af analysen er baseret på litteraturstudiet (bilag 1). På den baggrund peger analysen på, at der kan være forskellige måder at forstå dataetik:

- Dataetik kan forstås deskriptivt eller normativt
- Dataetik kan forstås inklusivt eller eksklusivt
- Dataetik kan forstås som et felt af problemstillinger eller som en praksis, der undersøger feltet

"Data ethics is a branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use."

Open Data Institute

Afslutningsvis formulerer analysen en definition af dataetik, som kan være nyttig for Rådets arbejde.

Hvad er en definition?

En definition er en beskrivelse af et begreb. Vi har alle ind imellem brug for definitioner, når vi støder på et ord, hvis betydning vi ikke kender. De fleste af os har også prøvet at formulere en definition, for eksempel når et barn gerne vil vide, hvordan man skal forstå "uretfærdigt", "kærlighed", eller "voksen".

En god definition er på en gang klar og præcis. Den fortæller på en enkel og letforståelig måde, hvad de centrale egenskaber er ved det begreb, som skal

defineres, og den trækker en tydelig grænse omkring begrebet, som viser, hvordan det adskiller sig fra andre begreber. En god definition af "dataetik" må derfor vise både, hvad det er, som karakteriserer dataetik, og hvordan det er anderledes end andre perspektiver på henholdsvis data og etik, eksempelvis "dataregulering" og "bioetik".

For at være klar må en definition også undgå at bruge det begreb, som skal defineres i selve definitionen. I definitionen af "dataetik" bør man derfor undgå at bruge begreberne "data" og "etik" eller i hvert fald præcisere, hvordan hvert af disse begreber skal forstås i netop denne sammenhæng.

I forskningssammenhænge formulerer man ofte en definition som et sæt af nødvendige og tilsammen tilstrækkelige betingelser. En nødvendig betingelse beskriver en egenskab, som et fænomen *skal* besidde for, at det kan være et eksempel på det begreb, som defineres. Dét at betingelserne tilsammen er tilstrækkelige betyder, at et fænomen, som har alle de egenskaber betingelserne peger på, *altid* er et eksempel på det begreb, som defineres.

Et enkelt eksempel på en definition, som består af nødvendige og tilsammen tilstrækkelige betingelser, er følgende:

> En retvinklet trekant er
> 1) en geometrisk figur
> 2) med tre sider og tre vinkler,
> 3) hvor netop én af vinklerne er 90 grader.

I slutningen af kapitlet formuleres en definition af dataetik, som har form af et sæt af nødvendige og tilsammen tilstrækkelige betingelser.

Leksikale og eksplikative definitioner

En god definition er både klar og præcis. Den er klar, hvis den på letforståelig vis karakteriserer og afgrænser begrebet. Men hvad vil det sige, at en definition er præcis? Svaret afhænger af formålet med definitionen.

Ofte forsøger en definition at indkredse, hvordan vi plejer at forstå et begreb. Sådanne definitioner kaldes leksikale. En leksikal definition er præcis, hvis den svarer til, hvordan personer normalt forstår begrebet.

En definition kan også foreslå en bestemt måde at skærpe vores forståelse af et begreb. Sådanne definitioner kaldes eksplikative. Definitionen kan eksempelvis inkludere en egenskab, selvom vi ikke typisk forstår denne egenskab som central for begrebet. Det kan skyldes, at der findes gode grunde til at mene, at det er vigtigt, om et fænomen har denne egenskab eller ej. En eksplikativ definition er præcis i den udstrækning, at definitionen karakteriserer og afgrænser begrebet på en måde, som er hensigtsmæssig i den relevante sammenhæng.

De to typer definitioner har hver deres fordele. Leksikale definitioner er nyttige til at slå fast, hvordan et sprogligt veletableret begreb anvendes. Eksplikative definitioner er nyttige, når vores forståelse af et begreb er uklar for eksempel fordi, der er tale om et nyt begreb, eller når en etableret forståelse overser en relevant forskel.

I det følgende arbejder vi både leksikalt og eksplikativt med definitionen af dataetik. Vi ser på, hvordan dataetik er blevet defineret i forsknings- og policy-sammenhænge. Vi vurderer også, hvordan eksisterende definitioner kan skærpes på en måde, som kan være nyttig for Dataetisk Råd.

Definitioner af dataetik

Dataetik er et begreb, som først de seneste år har tiltrukket sig væsentlig opmærksomhed. Af samme grund er begrebet fortsat sparsomt behandlet i den eksisterende litteratur.

Analysens litteraturstudie har identificeret tre eksplicitte definitioner i forskningslitteraturen og ni eksplicitte definitioner i policy-dokumenter. Af disse er syv kilder danske; de resterende er fra internationale forskningspublikationer og engelsksprogede policy-dokumenter.

> *"Data ethics is an emerging branch of applied ethics which describes the value judgements and approaches we make when generating, analyzing and disseminating data."*
>
> *UK Department for Digital, Culture, Media & Sport*

I tillæg til de eksplicitte definitioner diskuterer en lang række kilder data og etik på en måde, som implicit definerer dataetik. Litteraturstudiet har identificeret i alt 35 sådanne kilder. Visse af disse kilder arbejder eksplicit med definition af et nært beslægtet begreb, eksempelvis AI-etik og "big data"-etik.

Den måske mest oplagte definition af dataetik er, at det er den del af etikken, som angår data. Denne forståelse optræder centralt i et stort udsnit af kilderne. Den har endvidere den fordel, at den minder om etablerede, beslægtede definitioner, eksempelvis "dyreetik" og "straffeetik". En klar definition af dataetik må imidlertid præcisere, hvad "etik" og data" betyder i denne sammenhæng.

> *"Ethical issues are everywhere in the world of data, because data's collection, analysis, transmission and use can and often does profoundly impact the ability of individuals and groups to live well."*
>
> *Vallor 2018*

"Data" i dataetik

Det er i udgangspunktet enkelt at definere "data". Data betyder i denne sammenhæng digitalt lagret information. Imidlertid er det tydeligt i mange kilder, at dataetikkens fokus ikke er data som sådan, men snarere de *menneskelige handlinger* som vedrører data. Dataetikken omfatter eksempelvis skabelse, indsamling, lagring, processering, deling, analyse og anvendelse af data. Under ét kan vi sige, at dataetik omfatter de situationer, hvor en person *behandler* data.

> *"Ultimately, the central question is this: If big data is here to stay, in some sense, what kind of big data society do we want to have and how can we best achieve it?"*
>
> *Vayena & Tasioulas 2016*

En definition af dataetik, som omfatter alle situationer, hvor en person behandler data, er dog antageligt for bred. Mange måder at behandle data på er etisk trivielle og falder derfor ikke eller kun i begrænset omfang under dataetikken. Eksempler inkluderer bl.a. at tage et feriebillede med et digitalt kamera, at se en musikvideo på en hjemmeside eller at skrive en besked til en bekendt på sin telefon. Det er også

tydeligt i kilderne, at dataetikkens fokus typisk forstås som afgrænset til de situationer, hvor behandlingen af data stiller krav til agenten om at handle ansvarligt, forudsætter værdidomme, påvirker personers etiske interesser eller på anden vis, gør etikken central for handlingen. Vi kan sige, at dataetikkens fokus er handlinger, som behandler data, og hvor behandlingen af data i sig selv rejser en etisk problemstilling.

"Etik" i dataetik

Der er mindre klarhed over og større forskel på, hvordan kilderne forstår etik. Etik defineres blandt andet i) som et sæt af problemer, ii) som analyse af eller refleksion over problemerne, iii) som handlinger, der adresserer problemerne, iv) som bestemte måder at behandle data på, v) som afvejning af dilemmaer, vi) som principper for behandling af data, vii) som forståelse for afvejning af risici og fordele ved behandling af data og viii) som analyse af adfærd, som kan skade eller gavne personer.

Forskellene i, hvordan forskellige kilder forstår "etik", afspejler til dels forskelle mellem forfatternes etiske baggrundsteorier, eksempelvis mellem deontologiske og konsekvensetiske teorier. Disse forskelle er vigtige, men falder udenfor den nærværende kapitels analyse (se kapitel 2 om dataetiske principper og værdier). Sammenfattende kan vi sige, at sådanne etiske baggrundsteorier er teorier om, hvordan (etiske) værdier og principper påvirker, hvordan personer bør handle. Med det udgangspunkt kan vi sige, at behandling af data rejser en etisk problemstilling, når handlingen på en eller flere måder strider mod de værdier og principper, som påvirker, hvordan personer bør handle.

Det er vigtigt at være opmærksom på, at det er muligt for personer at være uenige om, hvorvidt en given situation er dataetisk relevant, *fordi* de er uenige om etiske baggrundsteorier. Dette behøver imidlertid ikke være udtryk for, at de er uenige om definitionen af dataetik som begreb. Begge parter kan acceptere, at dataetikkens fokus er handlinger, som behandler data, og hvor behandlingen af data i sig selv rejser en etisk problemstilling. Uenigheden skyldes i sådanne situationer, at personer, på grund af deres forskellige etiske baggrundsteorier, er uenige om, hvorvidt behandlingen af data i den pågældende situation i sig selv rejser en etisk problemstilling. En sådan uenighed er derfor ikke en udfordring for definitionen af dataetik, men alene for vurderingen af den mulige dataetiske problemstilling.

Tre forskellige måder at forstå dataetik

"Given the increase in the volume of personal data being collected and the use of automated methods to process these data for different purposes, one of the main priorities of the Data Ethics Commission is to reconcile the need to protect the individual's fundamental rights and freedoms – including self-determination and integrity – with the need to promote progress, prosperity, the safeguarding of democracy and the shaping of a society that is fit for the future."

German Dataethics Commission

Det er, som vi har diskuteret oven for, afgørende for definitionen af dataetik, hvordan man forstår henholdsvis "data" og "etik". Nogle tilsyneladende forskellige måder at forstå dataetik er imidlertid udtryk for forskelle i etiske baggrundsteorier, som ikke giver anledning til at definere dataetik på forskellige måder. Andre forskelle afspejler mere substantielt forskellige forståelser af, hvad etik og dataetik er.

I det følgende belyser vi tre centrale forskelle på, hvordan man kan forstå dataetik. Disse er:

- forskellen på at forstå etik deskriptivt og normativt
- forskellen på at forstå dataetik som felt og dataetik som praksis, og
- forskellen på at forstå dataetik inklusivt og eksklusivt.

På baggrund af de overvejelser, som vi gør os i den forbindelse, formulerer vi en definition af dataetik, som kan være nyttig for Dataetisk Råds arbejde.

Deskriptiv og normativ dataetik

Dataetik kan betyde forskellige ting. En afgørende forskel er, om "etik" i "dataetik" forstås deskriptivt eller normativt.

Når man forstår "etik" deskriptivt, så tænker man etik som det sæt af normer, der findes i en bestemt social sammenhæng. En sådan sammenhæng kunne eksempelvis være IT-branchen eller EU-

lande. Etik er i dette perspektiv et kulturelt og socialt fænomen, som kan studeres empirisk af eksempelvis sociologer og antropologer. Hvis dataetik forstås deskriptivt, så handler det derfor om de normer for behandling af data, som findes i en bestemt social sammenhæng, eksempelvis i Danmark.

Når man forstår "etik" normativt, så tænker man etik som det sæt af værdier og principper, som tilsammen afgør, hvordan vi bør handle. Normativ etik kan derfor i teorien være uafhængig af den sociale sammenhæng, en problemstilling optræder i. Hvis dataetik forstås normativt, så handler det derfor om de værdier og principper, som påvirker, hvordan vi bør behandle data.

Det er vigtigt at holde sig for øje, at deskriptiv og normativ etik er forskellige, men ikke modstridende forståelser af etik. En dataetisk problemstilling kan analyseres både fra et deskriptivt etisk perspektiv og et normativt etisk perspektiv. En deskriptiv analyse vil fokusere på, hvordan problemstillingen opfattes af en bestemt gruppe, og hvorfor gruppen forstår problemstillingen på netop denne måde. En normativ analyse fokuserer på hvilke værdier og principper, som er på spil i problemstillingen, og hvordan problemstillingen i lyset af disse bør løses. De to analyser anlægger forskellige perspektiver for at besvare forskellige spørgsmål. Begge typer analyse er meningsfulde, og når blot man er sig forskellen bevidst, kan de i nogle tilfælde supplere hinanden. Eksempelvis kan dét, at en bestemt måde at behandle data på krænker en udbredt norm, og at denne praksis derfor opleves som illegitim, i nogle tilfælde udgøre en normativ etisk begrundelse for ikke at behandle data på denne måde.

> *"I define a data ethics of power as an action-oriented analytical framework concerned with making visible the power relations embedded in the "Big Data Society" and the conditions of their negotiation and distribution, in order to point to design, business, policy, social and cultural processes that support a human-centric distribution of power."*
>
> *Hasselbach 2019*

Dataetik som felt og dataetik som praksis

En anden afgørende forskel er, om dataetik forstås som noget, personer undersøger eller som selve dét at undersøge. I nogle kilder forstås dataetik først og fremmest som et felt, eksempelvis som de problemstillinger, værdier og principper som knytter sig til behandlingen af data. I andre kilder forstås dataetik mere aktivt som en praksis, der undersøger eller vurderer dataetiske problemstillinger.

> *"A discussion of ethics and Big Data is dependent upon how one defines ethics. In general, ethics involves the analysis of conduct that can cause benefit or harm to other people… [Sound ethical theories enable the individual to make persuasive, logical and reasoned arguments based on the principles stated by the ethical theory."*
>
> *Herschel & Miori 2017*

Som med deskriptiv og normativ etik oven for er der tale om perspektiver, der kan supplere hinanden. Dataetik kan både forstås som et felt af problemstillinger og som en praksis med at analysere disse problemstillinger. I modsætning til normativ og deskriptiv etik ovenfor er der meget, der taler for, at en definition af dataetik bør integrere snarere end adskille de to perspektiver. Det vil sige, at dataetik bør forstås *både* som feltet af problemstillinger og som den praksis,der undersøger eller vurderer problemstillingerne.

Videnskabsteoretikere fremfører ofte med reference til den amerikanske videnskabshistoriker Thomas Kuhns arbejde, at såkaldt normalvidenskab udføres indenfor et paradigme, som består af en fælles forståelse af, hvad de relevante problemer er, hvordan disse problemer skal behandles og hvilke løsninger, man forsøger at producere.[1] Dataetik kan tilsvarende forstås som på en gang:

- et sæt af etiske problemstillinger, som knytter sig til behandlingen af data,

[1] Kuhn, T. S. (1996). The Structure of Scientific Revolutions. Chicago, University of Chicago Press.

- bestemte metoder til at arbejde systematisk med disse problemstillinger, eksempelvis vurdering af argumenter og
- en bestemt type løsninger, som arbejdet forsøger at producere, eksempelvis formuleringen af generaliserbare værdier og principper samt vurderingen af, hvordan disse værdier og principper påvirker konkrete etiske problemstillinger (se også analysen af dataetisk metode).

Afgrænsning af dataetikken

En tredje afgørende forskel knytter sig til, hvordan dataetik afgrænses i forhold til beslægtede felter af etikken.

Dataetik omfatter som felt problemstillinger, værdier og principper, som også findes i beslægtede felter, eksempelvis AI-etik, computer-etik, algoritme-etik, og Big Data-etik. En definition af dataetik må tage stilling til, hvordan grænserne mellem disse felter skal trækkes. Skal man adskille dataetikken, som et område der ikke behandler de problemstillinger, som findes i disse felter? Eller skal man tillade overlap, således at en problemstilling kan være dataetisk samtidig med, at den er en problemstilling i et af de andre felter?

> *"AI ethics is a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies."*
>
> *Alan Turing Institute*

En inklusiv definition af dataetik gør dataetik til så stort et felt som muligt. Den tillader problemstillinger at optræde i dataetik selvom, at de optræder eller endda er centrale i andre felter. Med en inklusiv definition kan hele felter, som AI-etik og Big Data-etik, vise sig at være underområder af dataetikken.

En eksklusiv definition begrænser dataetik som felt til at omfatte det sæt af problemstillinger, som ikke behandles i et eller flere andre felter. Derved defineres dataetik som et selvstændigt felt med et mere begrænset omfang.

Det er ikke klart, at en inklusiv definition stemmer bedre overens med vores etablerede begreb om dataetik end en eksklusiv definition, eller omvendt. Definitionen af dataetik er i kilderne ikke så skarpt afgrænset, at der kan siges at være en etableret betydning, som er mere eksklusiv eller mere inklusiv.

Valget af en inklusiv eller eksklusiv definition er derfor først og fremmest pragmatisk og bør afspejle Dataetisk Råds beslutning om, hvilke problemstillinger Rådet ønsker at arbejde med.

Sammenfatning og definition

Dataetik er et komplekst begreb. Både "data" og "etik"-delen af "dataetik" kan forstås på forskellige måder. Der findes ikke i de kilder, som analysens litteraturstudie har identificeret, en veletableret betydning, som forsøget på at definere begrebet kan trække på.

En klar og præcis definition kan imidlertid være med til både at afgrænse de problemstillinger, som Dataetisk Råd skal forholde sig til og til at angive, hvordan Rådet metodisk vil bearbejde disse problemstillinger. Af disse grunde kan det være hensigtsmæssigt for Rådet at antage en definition, som tager stilling til nogle af de forskellige måder, dataetik kan forstås på.

En sådan definition skal tage stilling til, hvilke typer teknologi og hvilke situationer med anvendelse af teknologi, som skal inkluderes i Rådets forståelse af "data". Den skal endvidere tage stilling til, om "etik" forstås deskriptivt eller normativt, som et felt, en praksis eller begge dele, og som knyttet til en bestemt etisk baggrundsteori eller mere skematisk.

> *"Data ethics can be defined as the branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing, and use), algorithms (including AI, artificial agents, machine learning, and robots), and corresponding practices (including responsible innovation, programming, hacking, and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values)."*
>
> *Floridi & Taddeo, 2016*

Et forslag til en definition af dataetik, som tager stilling til hvert af disse punkter, er følgende:

Dataetik er et sæt af problemstillinger, som vedrører hvordan personer bør handle, i situationer hvor handlingen involverer generation, indsamling, lagring, processering, analyse, deling eller anvendelse af data, og hvor denne denne behandling af data i sig selv udfordrer normative værdier eller principper. Dataetik er også den systematiske undersøgelse af dataetiske problemstillinger med henblik på at formulere velbegrundede svar på problemstillingerne, eksempelvis i form af generaliserbare principper for hvordan personer bør behandle data, eller i form af konkrete anbefalinger til afvejningen af konkurrerende hensyn i specifikke dilemmaer.

Litteratur om definitioner og begrebsanalyse

Burge, T. (1993). "Concepts, definitions, and meaning." Metaphilosophy 24(4): 309-325.

Gupta, A. (2019). Definitions. Stanford Encyclopedia of Philosophy. E. N. Zalta.

Hansson, S. O. (2010). "How to define: a tutorial." Princípios 13(19-20): 05-30.

Lippert-Rasmussen, K. (2016). Begrebsanalyse i politisk teori. Metode i normativ politisk teori. R. Sommer Hansen and S. Flinch Midtgaard. Viborg, Samfundslitteratur: 163-184.

Kapitel 2

Dataetiske værdier og principper

Dataetiske værdier og principper

Dataetisk Råd arbejder med at analysere og skabe debat om dataetik. I disse år oplever Danmark en hastig vækst i behandlingen af data, både i den private og den offentlige sektor.[2] I nogle situationer kan de nye teknologiske muligheder rejse dataetiske problemstillinger. I disse situationer kan vi kun træffe gennemtænkte valg ved grundigt at overveje, hvad der er på spil. Dataetisk Råds analyser og debat af dataetiske problemstillinger kan således udgøre et vitalt bidrag til, at Danmark kan være på forkant med den teknologiske udvikling. Men hvordan analyserer og diskuterer man dataetik?

Dataetik handler om det sæt af problemstillinger, hvor personers behandling af data i sig selv udfordrer etiske værdier og principper (se kapitel 1 om "Hvad er dataetik?"). Et afgørende skridt i overvejelsen af en dataetisk problemstilling er derfor at identificere og forstå de etiske værdier og principper, som er eller kan være på spil i dataetikken.

Der findes imidlertid ikke et etableret overblik over dataetiske værdier og principper. Dataetisk Råd har udarbejdet denne analyse for at afklare de værdier og principper, som ofte optræder i litteraturen om dataetiske problemstillinger, og redegøre for, hvordan man systematisk kan identificere relevante værdier og principper i arbejdet med dataetik.

Analysen kombinerer et litteraturstudie med perspektiveringer til relevante dele af den moderne forskning i etiske værdier (aksiologi) og principper (normativ teori).

Som forberedelse til præsentationen af dataetiske værdier og principper skitserer vi først, hvad henholdsvis en etisk værdi og et etisk princip er, og introducerer nogle distinktioner mellem forskellige typer værdier og principper. Der på redegør vi for, hvordan man i praksis kan identificere de etiske værdier og principper, som er på spil i en konkret dataetisk problemstilling. Endelig præsenterer vi et sæt af dataetiske værdier og principper, som vi i litteraturstudiet har set, spiller en prominent rolle i diskussionen af dataetik.

[2] I denne analyse anvender vi "behandling af data" som samlet betegnelse for de handlinger som på relevant vis vedrører data, f.eks. generation, indsamling, lagring, processering, analyse, anvendelse, og deling af data.

En central pointe i præsentationen af værdier og principper er, at begreberne ofte er flertydige og overlappende. Det betyder, at det i hver enkelt sammenhæng er afgørende at afklare, hvordan en værdi eller et princip skal forstås, samt hvilken etisk rolle det spiller.

Hvad er etiske værdier og principper?

Etik drejer sig om, hvordan personer bør handle. Dataetik drejer sig tilsvarende om, hvordan personer bør handle i netop de situationer, hvor databehandling i sig selv rejser etiske problemstillinger. Et oplagt spørgsmål er derfor: Hvilke forhold påvirker, hvordan personer bør handle?

Det er almindeligt i forskningen i etik at skelne mellem værdier og principper som to faktorer, der begge kan påvirke, hvordan personer bør handle.

- En værdi er noget, som gør verden bedre, når der er mere af det. Et almindeligt eksempel er menneskelig velfærd – det er alt andet lige bedre, at der er mere menneskelig velfærd, end at der er mindre menneskelig velfærd. Værdier handler om, hvornår noget er etisk godt eller dårligt.
- Et princip er en generel etisk grund til at handle på en bestemt måde. Et almindeligt eksempel er skadesprincippet – personer har en (stærk) etisk grund til ikke at handle på måder, som skader andre. Principper handler om, hvornår en handling er etisk rigtig eller forkert.

Dataetiske principper er afgørende for dataetikken, fordi det er principperne som på et generelt niveau beskriver, hvilke grunde personer har til at handle. Dataetiske værdier er vigtige, fordi mange principper handler om værdier. Eksempelvis er et plausibelt bud på, hvad det vil sige at skade en person, at det betyder at reducere vedkommendes velfærd. Velfærd er en etisk værdi. Sådanne principper får altså kun indhold, når vi kombinerer dem med de relevante værdier. Det betyder også, at det samme princip kan pege på meget forskellige handlinger afhængigt af, hvordan man forstår de værdier, som det kombineres med.

Både værdier og principper kan videre inddeles i to typer: intrinsiske og instrumentelle værdier og grundlæggende og mellemniveau-principper. Nedenfor skitserer vi forskellen på disse, hvorpå vi diskuterer, hvordan man kan identificere dataetiske værdier og principper i en dataetisk problemstilling.

Intrinsiske og instrumentelle værdier

En værdi er noget, som gør verden bedre, når der er mere af det. Det er imidlertid vigtigt at skelne mellem ting, som kun er instrumentelt værdifulde og ting, som er intrinsisk værdifulde.

En instrumentel værdi er noget, som kun gør verden bedre, fordi det fører til, at der bliver mere af en anden værdi. Et almindeligt eksempel er penge, som alene har værdi, fordi de kan veksles til andre ting, som er værdifulde.

En intrinsisk værdi er noget, som *i sig selv* gør verden bedre, når der er mere af det. Vi har ovenfor nævnt menneskelig velfærd som et almindeligt eksempel. Velfærd er næppe værdifuldt, *fordi* det fører til andre værdier. Velfærd er ganske enkelt etisk godt i sig selv.

Når man forsøger at identificere og vurdere dataetiske værdier, er det vigtigt at holde sig forskellen mellem instrumentelle og intrinsiske værdier for øje.

Grundlæggende og mellemniveauprincipper

Vi har oven for vist, hvordan man bør skelne mellem intrinsiske og instrumentelle værdier. Der findes en beslægtet forskel på to typer etiske principper: grundlæggende principper og mellemniveauprincipper.

Grundlæggende principper er ikke begrundet i andre principper. Et grundlæggende princip er, som navnet antyder, etikkens fundament. Sådanne principper er typisk universelle: de gælder i alle sammenhænge, uanset hvilke andre grunde der måtte være på spil.

Mellem-niveauprincipper er en slags etiske tommelfingerregler. De er på den ene side ikke en anvisning om, hvordan vi bør handle i en bestemt situation. Der er fortsat tale om et forholdsvist generelt princip. På den anden side er de nærmere på de konkrete problemstillinger, hvor de skal anvendes, end et grundlæggende princip.

Mange af de etiske principper, som vi diskuterer og anvender i vores hverdag, er mellem-niveauprincipper. Et eksempel er bl.a., at "det er forkert at lyve". Når voksne lærer børn princippet, er det fordi, det er let at anvende og giver det rigtige svar på, hvordan vi bør handle i de fleste almindelige sammenhænge. De fleste anerkender imidlertid også, at der findes "hvide løgne", det vil sige situationer, hvor det ikke er forkert at lyve eller måske endda forkert ikke at lyve. Eksempelvis

er det nok de færreste, som vil hævde, at det er etisk forkert at føre fødselaren bag lyset for at planlægge og afholde en overraskelsesfest.

Mellem-niveauprincipper har to fordele. For det første er de typisk let anvendelige retningslinjer, som gør det muligt hurtigt og enkelt at træffe etiske beslutninger. Samtidig kan mellemniveauprincipper i nogle tilfælde være mindre kontroversielle end grundlæggende principper. Det forekommer, når flere forskellige grundlæggende principper alle kan støtte det samme mellem-niveauprincip.[3] I sidste ende er et mellem-niveauprincip imidlertid kun en tilnærmelse til grundlæggende principper. Det betyder, at et mellem-niveauprincip kan være upræcist, især når det anvendes på usædvanlige situationer.

Ligesom med forskellen mellem intrinsiske og instrumentelle værdier er det vigtigt, når man forsøger at identificere og vurdere dataetiske principper, at skelne mellem, om et princip er grundlæggende eller et mellem-niveauprincip.

Identifikation og analyse af værdier og principper

Oven for har vi introduceret, hvad henholdsvis en værdi og et princip er og skelnet mellem to forskellige typer af hver. Sådanne værdier og principper er centrale i dataetikken, fordi de tilsammen afgør, hvordan man etisk set bør handle. Når man skal analysere en konkret dataetisk problemstilling, er det derfor afgørende at identificere de værdier og principper, som er på spil. Hvordan gør man det?

Én fremgangsmåde, som umiddelbart kan virke tiltalende, er at definere et sæt af universelle dataetiske værdier og principper og derpå undersøge, hvilke af dem, som berøres i hver aktuel problemstilling.

Denne tilgang møder imidlertid to alvorlige udfordringer. Den første er, at det kan være vanskeligt at vurdere, om en etisk værdi eller et etisk princip er berørt i en konkret problemstilling. Det skyldes, at universelle værdier og principper nødvendigvis vil være formuleret både abstrakt og generelt. Meget ofte har de værdier og principper, som man møder i eksempelvis den offentlige debat om

[3] Et berømt eksempel er den australske moralfilosof Peter Singers princip, at "hvis man kan forhindre noget moralsk meget dårligt, og det kun har en meget lille omkostning for en selv at forhindre det, så er man moralsk forpligtet til at gøre det". Se Singer, P. (1972). "Famine, Affluence, and Morality." *Philosophy & Public Affairs* **1**(3): 229-243.

dataetiske problemstillinger, da også en mere specifik karakter, som medfører, at de kun tentativt kan fortolkes som varianter af en intrinsisk værdi eller et grundlæggende princip.

Den anden udfordring er, at det kan være vanskeligt at definere et ukontroversielt sæt af universelle dataetiske værdier og principper. Der findes således ikke i hverken den etiske forskning mere bredt eller i den specifikt dataetiske forskning konsensus om et sæt af etiske værdier og principper.

En alternativ fremgangsmåde er at tage udgangspunkt i hver enkelt dataetiske problemstilling (se kapitel 4 om "Dataetisk metode og teori"). I en konkret dataetisk problemstilling kan de relevante dataetiske værdier og principper identificeres ved at isolere og analysere *begrundelser* for etiske synspunkter.

Etiske synspunkter og begrundelser

Et etisk synspunkt på en dataetisk problemstilling er en konklusion om, hvordan man bør handle, for eksempel at en bestemt måde at behandle data på er etisk forkert, og at vi derfor bør lade være med at behandle data på denne måde.

Når vi vil forsvare et synspunkt, så giver vi begrundelser. Almindelige synspunkter om, hvordan verden er indrettet, understøttes af *deskriptive* begrundelser, som henviser til, hvordan verden *er*. Hvis vi eksempelvis vil forsvare det synspunkt, at globale klimaforandringer er menneskeskabte, så kan vi henvise til, at mennesker er ansvarlige for at udlede drivhusgasser, og at et øget niveau af drivhusgasser i atmosfæren skaber klimaforandringer.

Hvis vi vil forsvare et etisk synspunkt, så giver vi også begrundelser. Etiske synspunkter har imidlertid den særlige karakter, at de ikke kan forsvares alene med deskriptive begrundelser. Hvis vi vil forsvare et etisk synspunkt, så er vi før eller siden nødt til at sige noget om, hvordan verden *bør være*.[4] Det vil sige, at vi giver *etiske* begrundelser. En etisk begrundelse trækker eksplicit eller implicit på etiske værdier og principper. Ved at analysere de etiske begrundelser for synspunkter på den dataetiske problemstilling kan man således identificere etiske værdier og principper, som *potentielt* er berørt i problemstillingen.

[4] Forsøg på at forsvare etiske synspunkter med deskriptive begrundelser alene kaldes for "den naturalistiske fejlslutning".

Arbejdet med at identificere berørte etiske værdier og principper rejser imidlertid to yderligere opgaver. Den første opgave handler om, at der kan være etiske begrundelser, som ved nærmere eftersyn viser sig at være uplausible. Det kan skyldes, at den etiske værdi eller det etiske princip, som begrundelsen trækker på, må afvises. Ikke alle etiske begrundelser trækker på værdier og principper, som vi har grund til at acceptere. Det kan også skyldes, at det viser sig, at det princip eller den værdi, som begrundelsen hævder er på spil, faktisk ikke er på spil. I sådanne tilfælde trækker begrundelsen på et plausibelt princip eller en plausibel værdi, men tager fejl ved at hævde, at problemstillingen berører det pågældende princip eller den pågældende værdi. For hver potentielt relevant dataetisk værdi eller princip er man derfor nødt til kritisk at vurdere, om vi bør acceptere værdien eller princippet, samt om værdien eller princippet faktisk er berørt af problemstillingen.

Den anden komplikation er, at der kan være oversete synspunkter og begrundelser. Dataetik er et nyt felt, og man må derfor forvente, at mange dataetiske problemstillinger kun er blevet helt overfladisk behandlet både i forskningen og den offentlige debat. Det betyder, at der kan være plausible synspunkter og begrundelser, som ikke er blevet fremført. Hvis man overser disse, så overser man tilsvarende nogle af de værdier og principper, som berøres af problemstillingen. I disse tilfælde er løsningen er at forsøge at formulere de oversete synspunkter og begrundelser, som plausibelt kunne fremføres i tilknytning til problemstillingen. En sådan kreativ formulering af mulige, plausible synspunkter og begrundelser er i sagens natur vanskelig, og der findes derfor uundgåeligt en risiko for, at man ikke får identificeret alle relevante værdier og principper.

Processen med at identificere relevante værdier og principper i en konkret dataetisk problemstilling kan samlet opsummeres i de følgende tre skridt:

1) Identificer etiske synspunkter på den dataetiske problemstilling.
2) Identificer begrundelser for de etiske synspunkter på problemstillingen.
3) Afklar og evaluer de etiske værdier og principper, som er på spil i begrundelserne.

Værdier og principper i dataetikken

En væsentlig opgave for denne analyse har været at identificere og kategorisere dataetiske værdier og principper i litteraturen om dataetik (se bilag 1). I dette afsnit introducerer vi nogle af de væsentligste værdier og principper, som optræder i denne litteratur:

- Ansvarlighed
- Autonomi
- Demokrati
- Etik
- Fairness
- Gennemsigtighed
- Godgørenhed
- Lighed
- Privatliv
- Retfærdighed
- Sikkerhed
- Velfærd
- Værdighed

For hver værdi eller princip afklarer vi, hvordan det kan forstås i dataetisk sammenhæng og henviser til beslægtede værdier og principper.

En gennemgående pointe er, at mange værdier og principper er komplekse eller flertydige begreber. For at analysere en dataetisk problemstilling i lyset af en værdi eller et princip er det derfor afgørende, at man gør sig klart *præcis,* hvad det pågældende begreb betyder i netop den givne sammenhæng, herunder hvilken etisk rolle det antages at spille.

Ansvarlighed

Det engelske "accountability" kan på dansk oversættes til "ansvarlighed". Ansvarlighed i dataetik handler om at sikre, at der findes agenter, som kan stilles til ansvar for behandlingen af data.[5]

Et vigtigt spørgsmål i diskussionen af ansvarlighed er, hvad det vil sige, at en agent (en person, myndighed eller virksomhed) er moralsk ansvarlig for databehandling eller konsekvenserne af databehandling. Der findes i forskningslitteraturen en kompleks diskussion af, hvad det vil sige, at en agent er moralsk ansvarlig.[6]

På et meget generelt niveau fremhæves ofte tre betingelser for, at en agent er moralsk ansvarlig for en bestemt situation:

1) Situationen er en kausal konsekvens af agentens handling,
2) agenten havde relevant viden om, at situationen kunne følge af handlingen og

[5] Se især Nissenbaum, H. (1996). "Accountability in a computerized society." *Science and Engineering Ethics* **2**(1): 25-42; Kroll, J., et al. (2017). "Accountable Algorithms." *University of Pennsylvania Law Review* **165**(3): 633; MSI-AUT (2018). "A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework", *Council of Europe*; New, J. and D. Castro (2018). How Policymakers Can Foster Algorithmic Accountability, *Center for Data Innovation*. Et udmærket overblik findes i Noorman, M. (2020). "Computing and Moral Responsibility." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

[6] For et overblik, se Talbert, M. (2019). "Moral Responsibility." I: E. N. Zalta (Ed.). *Stanford Encyclopedia of Philosophy*. Bemærk, at det er et åbent spørgsmål i forskningen, om menneskelige agenter kan indfri betingelserne for moralsk ansvar. Se eksempelvis Levy, N. (2011). *Hard Luck*. Oxford, Oxford University Press. Pereboom, D. (2001). *Living without Free Will*. Cambridge, Cambridge University Press; og på dansk, Lippert-Rasmussen, K. (1999). *Viljens frihed og moralsk ansvar*. København, Nyt Nordisk Forlag.

3) agenten havde kontrol over handlingen, for eksempel ved at agenten havde muligheden for at handle anderledes.

Ansvarlighed fremhæves ofte i dataetikken, fordi databehandling i nogle situationer fører til, at ansvaret for visse konsekvenser bliver mere diffust. Hvis eksempelvis en offentlig myndighed fejlagtigt afviser en borgers ansøgning om at modtage en offentlig ydelse, så vil det ofte være enkelt at pege på en agent, som er ansvarlig for fejlen. Der kan for eksempel være en sagsbehandler, som traf beslutningen, og derfor er kausalt ansvarlig for afvisningen (1), som vidste at dette ville have den konsekvens, at borgeren ikke modtog ydelsen (2) og som havde mulighed for at handle anderledes ved at godkende ansøgningen (3).

Hvis myndigheden erstatter menneskelige sagsbehandlere med en software-algoritme, som træffer automatiserede afgørelser, så kan det blive vanskeligt at fastslå, hvor ansvaret for fejl skal placeres. Er det for eksempel den myndighed, som har indkøbt og implementeret den automatiserede beslutningsstøtte, som er ansvarlig? Er det den myndighed, som anvender den automatiserede beslutningsstøtte og udmønter beslutningen? Eller er det udvikleren af software-algoritmen? Det vil i nogle tilfælde være vanskeligt at finde én agent, som på samme tid indfrier alle tre betingelser for at være moralsk ansvarlig for, at borgerens ansøgning er blevet afvist.

En nært beslægtet problematik drejer sig om, hvilke agenter vi *bør holde* ansvarlige for en given måde at behandle data på eller for konsekvenserne af denne databehandling. Svaret på dette spørgsmål behøver ikke at være det samme som på spørgsmålet om, hvilken agent der *er* ansvarlig. Det kan eksempelvis være tilfældet, at det er vanskeligt entydigt at placere ansvaret for en bestemt databehandling på én agent, men ikke desto mindre tilfældet, at vi har gode grunde til at *holde* en bestemt agent ansvarlig, for eksempel fordi det vil minimere antallet af fejl, hvis netop denne agent holdes ansvarlig.

Ansvarlighed i dataetikken kan således betyde både, at databehandlingen organiseres, så det er klart hvilken agent, der er ansvarlig for databehandlingen eller dens konsekvenser, og at det gøres klart hvilke agenter, vil blive holdt ansvarlige for databehandlingen.[7]

[7] En tredje betydning af ansvar handler om at "opføre sig ansvarligt". Det betyder i denne sammenhæng at reflektere over dataetik, og handle i overensstemmelse med resultatet af refleksionerne. Se "Etik" nedenfor.

Autonomi

Det engelske "autonomy" hedder på dansk "autonomi". Autonomi betyder i etisk sammenhæng en persons evne til at træffe meningsfulde beslutninger om, hvordan vedkommende vil leve sit liv.

Autonomi i dataetikken kan handle om at sikre, at databehandlingen ikke begrænser de berørte personers mulighed for at træffe meningsfulde valg om deres eget liv. Mange etiske teorier hævder imidlertid også, at personers autonomi danner grundlag for mere generelle etiske principper. Autonomi i dataetikken kan derfor også handle om, at databehandlingen respekterer generelle principper baseret på personlig autonomi.

Autonomi er et centralt begreb i store dele af den etiske forskning. Begrebet autonomi kan imidlertid fortolkes på flere forskellige måder, og det spiller væsentligt forskellige roller i forskellige sammenhænge, eksempelvis i diskussionerne af moralsk ansvar, generelle etiske teorier, personlig frihed og paternalisme.[8] Det er derfor vigtigt at gøre sig klart, hvilken betydning af autonomi, der er på spil i hver enkelt konkret problemstilling.

På et meget generelt niveau kan man skelne mellem autonomi i en bred og i en snæver forstand, samt mellem autonomi som værdi og autonomi som grundlag for etiske principper.

I en snæver forstand handler autonomi om en persons evne til at træffe meningsfulde valg om, hvordan vedkommende vil leve sit liv. Et meningsfuldt valg betyder i denne sammenhæng et valg, som er funderet i personens *autentiske* grunde og værdier. Modsætningen til et autonomt valg i denne betydning er et valg, som er truffet under indflydelse af manipulation, uvidenhed eller værdier, som personen ikke reelt identificerer sig med.

I en bredere forstand handler autonomi også om, hvilke muligheder en person kan realisere på baggrund af en autonom beslutning. Hvor autonomi i snæver forstand knytter sig til de indre eller

[8] Et godt overblik findes i Christman, J. (2015). "Autonomy in Moral and Political Philosophy." I: E. N. Zalta (Ed.). *Stanford Encyclopedia of Philosophy*.

psykologiske barrierer for at træffe meningsfulde valg, så knytter autonomi i denne bredere forstand sig til de eksterne (fysiske, retslige, sociale) barrierer for personers evne til at træffe meningsfulde valg om deres liv.

Eksempelvis vil en kvinde, som lever i et stærkt konservativt, patriarkalsk samfund kunne være autonom i den snævre betydning, selvom hun er underlagt retslige og sociale barrierer, som begrænser hendes muligheder for at leve sådan, som hun ønsker. En anden kvinde i det samme samfund, som internaliserer negative stereotyper om kvinder, kan derimod miste autonomi også i den snævre betydning, hvis hendes vilje derved underlægges værdier og forestillinger, som hun dybest set ikke identificerer sig med.

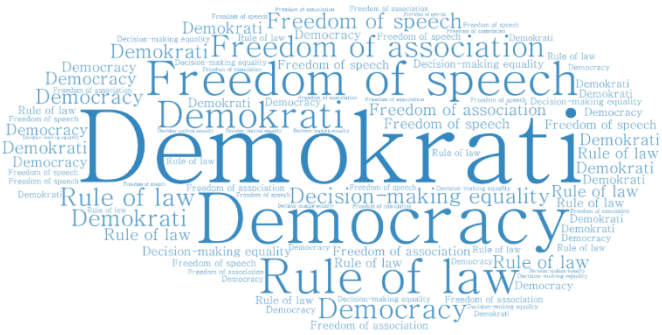
Hvis autonomi forstås som en værdi, så vil den i dataetikken være relevant i de situationer, hvor databehandling enten medfører, at det bliver sværere for en person at træffe beslutninger i overensstemmelse med sine autentiske grunde og værdier, eller når en person får sværere ved at realisere visse muligheder, som vedkommende kunne ønske sig at forfølge på baggrund af et autonomt valg.

Når autonomi forstås som kernen i et etisk princip, kan det begrunde eksempelvis moralske restriktioner. En autonomi-baseret moralsk restriktion kan for eksempel være et princip som hævder, at vi har stærke etiske grunde til ikke at udføre bestemte typer handlinger, fordi det ville udvise manglende respekt for personers autonomi at handle på denne måde.

Demokrati

Flere tekster fremhæver demokrati, demokratiske institutioner og socio-kulturelle forudsætninger for demokrati som dataetisk relevante værdier.

Demokrati kan forstås både i en bredere og en mere snæver betydning. I en snæver betydning er demokrati blot en måde at træffe kollektive beslutninger på, som er karakteriseret ved, at alle deltagere i bestemte henseender har lige indflydelse på beslutningen.[9] I en bredere forstand inkluderer demokrati også visse af de institutioner og rettigheder, som vi forbinder med demokratiske lande. Det kan eksempelvis være retssikkerhed, ytringsfrihed og beskyttelse af sårbare minoriteter mod diskrimination[10]. Selvom sådanne institutioner og værdier ofte bliver kategoriseret som demokratiske værdier, kan det være en fordel i arbejdet med en dataetisk problemstilling at præcisere, hvilke dele af demokratiet (i bred forstand) der er på spil.



Et selvstændigt spørgsmål angår, hvordan demokrati – i bred eller snæver forstand – er etisk relevant. Dette spørgsmål behandles normalt under overskriften "normativ demokratiteori". Demokrati italesættes ofte som intrinsisk værdifuldt, men det er også muligt at forklare dets værdi instrumentelt, eksempelvis ved at henvise til at demokrati har tendens til at sikre et højt niveau af menneskelig velfærd, eller ved at det har tendens til at maksimere menneskelig autonomi. Demokrati kan også have etisk relevans, hvis der findes etiske principper, som fordrer demokrati. Endelig kan de tre muligheder kombineres på forskellige måder, eksempelvis ved at man antager både at demokrati har instrumentel værdi, og at der findes etiske principper, som fordrer demokrati.

For at kunne vurdere hvilken rolle hensynet til demokrati bør spille i en dataetisk problemstilling, er man således nødt til at præcisere både begrebet om demokrati og hvilken etisk rolle, det antages at spille.

[9] For et godt overblik over både debatten om definitionen af demokrati og normative demokratiteori, se Christiano, T. (2006). "Democracy." I: E. N. Zalta (Ed.). *Stanford Encyclopedia of Philosophy.*

[10] Se Held, D. (2006). *Models of Democracy*. Cambridge, Polity Press; Dahl, R. A. (2000). *On Democracy.* New Haven, Yale Nota Bene.

Etik

En række kilder fremhæver etik i sig selv som en værdi, der har betydning for dataetik og databehandling.

Etik skal i denne sammenhæng forstås som bevidsthed om dataetiske problemstillinger, systematiske refleksioner over dem og motivation til at handle i overensstemmelse med, hvad sobre dataetiske overvejelser måtte nå frem til. Det hævdes eksempelvis, at det er afgørende, at databehandling sker under behørig hensyntagen til de personer, den berører (human-centrism), at relevante overvejelser og hensyn til etiske værdier integreres i databehandlingen (value-by-design) eller at bevidsthed om dataetiske problemstillinger, og hvordan de løses, er afgørende hos de agenter, der udfører databehandling (ethical awareness). Dataetisk bevidsthed diskuteres også i nogle sammenhænge i tilknytning til ansvar og ansvarlighed, når betydningen er at *tage* ansvar for at *handle* ansvarligt.

Én plausibel måde at forstå, hvordan dataetisk bevidsthed er relevant for dataetikken, er, at den er instrumentelt værdifuld. Det kan eksempelvis være fordi, den er en forudsætning for eller fremmer, at databehandlingen udføres i overensstemmelse med (andre) dataetiske værdier og principper. Især i de situationer, hvor databehandling rejser komplekse dataetiske problemstillinger, vil det typisk give en agent de bedste forudsætninger for at handle etisk forsvarligt, hvis agenten er dataetisk bevidst.

Visse deontologiske teorier hævder endvidere, at det er afgørende for en handling, hvilken intention agenten har. Ifølge sådanne teorier påvirker agentens subjektive *grunde* til at handle, om handlingen er rigtig eller forkert. Hvis intentioner spiller en sådan rolle, kan der således også være principper, som kræver, at agenter, som udfører databehandling, er bevidste om dataetik.

Fairness

Det engelske "fairness" bruges i dag også p dansk. I dataetikken knytter diskussionen af fairness sig især til spørgsmålet om, hvordan data om særligt beskyttede karakteristika, som køn, etnicitet, seksualitet, religion og handicap, kan anvendes i databehandlingen, samt om hvordan databehandling kan have forskellig effekt for forskellige grupper.[11]

Et eksempel kan være en algoritmisk model, som tolker røntgenbilleder for at identificere ondartede kræftknuder, og som har lavere fejlrate, når den vurderer mænd, end når den vurderer kvinder. I sådanne tilfælde kan et princip, som kræver fairness, måske fordre, at den algoritmiske model justeres, således at den har mere ensartet fejlrate for de to grupper.

Fairness forstås typisk som et princip knyttet til retfærdighed (se nedenfor), som fordrer, at personer stilles lige i bestemte henseender. Fairness kan eksempelvis fordre, at personer *behandles* lige, eller at de sikres *lige muligheder*. Fairness kan derfor også forstås som krav om upartiskhed, procedural lighed og ikke-diskrimination, herunder at bestemte forhold såsom køn og etnicitet ikke påvirker, hvor godt eller dårligt personer er stillet.[12] Derimod fordrer fairness typisk ikke substantiel lighed, det vil sige, at personer stilles lige i den endelige fordeling af goder, omend en mere lige fordeling kan være en forudsigelig konsekvens af mere lige muligheder.

Det er i diskussionen af fairness i dataetik afgørende både at afklare, præcis, hvad det vil sige at behandle personer fair, det vil sige i hvilken forstand, de skal være stillet lige, og hvor stærkt hensynet til fairness er i sammenligning med andre hensyn. En vurdering af, hvilken etisk vægt fairness skal

[11] Se Chouldechova, A. and A. Roth (2018) "The Frontiers of Fairness in Machine Learning." *arXiv e-prints*; Binns, R. (2018). "Fairness in Machine Learning: Lessons from Political Philosophy." *Journal of Machine Learning Research* **81**: 1-11; Berk, R., H. Heidari, S. Jabbari, M. Kearns and A. Roth (2018). "Fairness in Criminal Justice Risk Assessments: The State of the Art." *Sociological Methods & Research* **Online first**.

[12] For et godt overblik, se Arneson, R. (2015). "Equality of Opportunity." I: E. N. Zalta. *Stanford Encyclopedia of Philosophy*.

gives, er afgørende, fordi der ofte vil være situationer, hvor hensynet til fairness trækker i én retning, mens andre etiske værdier og principper trækker i en anden.

I eksemplet med en algoritmisk model, som identificerer kræftknuder, kan det for eksempel være sådan, at vi kun kan stille de to grupper mere lige på bekostning af modellens overordnede præcision. Det betyder, at mere fairness vil have en omkostning i form af flere fejl. I sådanne situationer er en dataetisk analyse nødt til at vurdere, hvordan man skal afveje de to konkurrerende hensyn.

Gennemsigtighed

Det engelske "transparency" kan på dansk oversættes til "gennemsigtighed". Gennemsigtighed er i dataetikken en værdi, som handler om, at databehandlingen er forståelig både for de agenter, som udfører databehandling, for de personer, som er berørt af databehandling og for agenter, som skal overvåge og administrere databehandling (for eksempel myndigheder).[13]

Gennemsigtighed i databehandling udfordres grundlæggende af tre forhold:

- Begrænsninger på adgang til information om databehandlingen
- Tekniske forudsætninger for forståelse af databehandlingen
- Uoverskuelig kompleksitet i databehandlingen[14]

Begrænsninger på adgang til information om databehandlingen skyldes normalt, at de agenter, som udfører databehandling, har interesse i at begrænse adgang til visse informationer

[13] Se Zerilli, J., A. Knott, J. Maclaurin and C. Gavaghan (2019). "Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?" *Philosophy & Technology* **32**: 661-683; Veale, M., M. Van Kleek and R. Binns (2018). "Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*: 1-14; Lepri, B., N. Oliver, E. Letouzé, A. Pentland and P. Vinck (2018). "Fair, Transparent, and Accountable Algorithmic Decision-making Processes." *Philosophy & Technology* **31**(4): 611-627.

[14] Se Molnar, C. (2020). *Interpretable machine learning. A guide for making black box models explainable*; Rudin, C. (2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead." *Nature Machine Intelligence* **1**(5): 206-215.

om databehandlingen. Denne interesse kan eksempelvis skyldes, at disse informationer udgør forretningshemmeligheder, eller at adgang til informationerne kan påvirke databehandlingens kvalitet, for eksempel fordi personer ændrer adfærd, hvis de får adgang til information om, hvordan databehandlingen fungerer.

Tekniske forudsætninger for forståelse af databehandlingen er et grundvilkår ved mange typer databehandling, fordi teknologien er relativt kompleks og trækker på et specialiseret ordforråd. Det betyder, at det kan være vanskeligt for lægpersoner at forstå databehandlingen, selv med adgang til den relevante information om den.

Endelig er visse former for databehandling så kompleks, at det kan være vanskeligt selv for eksperter at forstå, hvordan databehandlingen foregår. De mest omdiskuterede eksempler er brugen af maskinlæring til at træne avancerede algoritmer, for eksempel dybe neurale netværk, som kan analysere data. I nogle tilfælde består sådanne algoritmer af så mange enkelte komponenter, som gensidigt påvirker hinanden, at det er umuligt for mennesker at overskue, hvordan algoritmen samlet set behandler data.

Et vigtigt spørgsmål, som fortsat er underbelyst i forskningen, er, hvordan gennemsigtighed er etisk relevant. Et plausibelt bud er, at gennemsigtighed er instrumentelt værdifuldt, eksempelvis fordi det fremmer tillid til databehandlingen, sikkerhed og kontrol med data. En anden mulighed er, at der findes etiske principper knyttet til gennemsigtighed, eksempelvis et princip som fordrer, at personer har adgang til indsigt i, hvordan deres data behandles for at beskytte deres autonomi. Det er indlysende afgørende for at kunne behandle gennemsigtighed i en konkret dataetisk problemstilling, at man afklarer hvilken type gennemsigtighed, som er på spil, om gennemsigtighed er en dataetisk værdi eller et princip samt præcis hvilken etiske betydning, det har.

Godgørenhed

Det engelske "beneficence" kan på dansk oversættes til "godgørenhed". Godgørenhed er et etisk princip, som fordrer, at man handler på en måde, som fremmer det gode.[15]

[15] For et godt overblik, se Beauchamp, T. (2019). "The Principle of Beneficence in Applied Ethics." I: E. N. Zalta (Ed.). *Stanford Encyclopedia of Philosophy*.

For at vurdere om en handling fremmer det gode, kigger man på handlingens konsekvenser og evaluerer, hvilken værdi disse konsekvenser etisk set har. En enkel metode er at spørge, om konsekvenserne af handlingen gør verden bedre eller dårligere målt som summen af etiske værdier.

Princippet om godgørenhed kræver derfor, at man på forhånd har afgjort, hvilke etiske værdier der skal bruges til at evaluere handlingen. Den samme handling kan vurderes meget forskelligt afhængigt af, hvilke etiske værdier man antager.

Godgørenhed kan spille en rolle i mange dataetiske problemstillinger, eksempelvis når databehandling har utilsigtede, men skadelige konsekvenser. Princippet om godgørenhed forklarer det intuitive synspunkt, at dette udgør et etisk problem ved databehandlingen.

En vigtig pointe er, at princippet om godgørenhed antager, at der er lige så stærke grunde til at handle på en måde, som gør verden bedre (gavn), som der er til *ikke* at handle på en måde, som gør verden dårligere (skade). En person, som har mulighed for at gøre verden bedre, men ikke gør det, handler derfor lige så etisk forkert som en person, der handler på en måde, der gør verden værre. Princippet om godgørenhed er også krævende, fordi det antager, at vi har grund til at fremme det gode *mest muligt*.

Der findes imidlertid beslægtede principper, eksempelvis skadesprincippet og proportionalitetsprincipper, som er mindre krævende. Et skadesprincip fokuserer alene på, at agenter ikke bør handle på en måde, som gør verden dårligere (skader). Et proportionalitetsprincip hævder, at en handling, som i én forstand har dårlige konsekvenser, skal have *tilstrækkeligt* gode konsekvenser i andre henseender, men ikke at vi i almindelighed har grund til at gøre verden bedre.

Lighed

Lighed i dataetik handler om at sikre, at personer er substantielt lige. Når lighed spiller en etisk rolle, fordrer det således, at vi fordeler goder så ligeligt som muligt.[16]

Det vigtigste spørgsmål, når man skal afklare, hvilken etisk rolle lighed spiller, er, hvilke goder der skal fordeles ligeligt. [17] Det kan eksempelvis være økonomiske goder, men også rettigheder, status, positioner i samfundet eller menneskelig velfærd. Det er afgørende, hvilket gode man mener, at vi har etiske grunde til at fordele ligeligt, fordi et samfund, som er lige med hensyn til fordelingen af ét gode, kan være meget ulige med hensyn til fordelingen af andre goder.

Mindst to yderligere spørgsmål er vigtige i afklaringen af, hvilke rolle lighed spiller. For det første om vi alene har grund til *ikke* at handle på måder, som skaber mere ulighed, eller om vi har grund til at tilstræbe mere lighed. I det andet tilfælde kan lighed forstås som en værdi, således at én situation er bedre end en anden, hvis den er mere lige. I det første tilfælde er der snarere tale om et princip, som kræver, at personer ikke øger eksisterende ulighed.

For det andet er man nødt til at afklare, hvordan hensynet til lighed skal afvejes imod andre principper. Eksempelvis vil det ofte være muligt at skabe mere lighed alene ved at sørge for, at de bedst stillede får færre goder. En sådan handling, som ikke stiller nogen bedre, vil være etisk problematisk ifølge for eksempel et princip om godgørenhed.[18] Hvordan skal vi i en sådan situation vurdere, hvor tungt henholdsvis lighed og godgørenhed vejer?

---

[16] Et godt overblik findes i Arneson, R. J. (2013). "Egalitarianism." I: E. N. Zalta (Ed.). *Stanford Encyclopedia of Philosophy*.

[17] Et bud, som vi har behandlet ovenfor, er muligheder. Et princip om fairness fortolkes ofte netop som et krav om, at personer får de samme muligheder. Fairness og mulighedslighed kan imidlertid med fordel behandles selvstændigt. Vi begrænser derfor substantiel lighed til at handle om den lige fordeling af andre goder.

[18] I forskningen diskuteres sådanne eksempler som indvendinger mod lighed (egalitarisme) under betegnelsen "levelling down objections". Se Parfit, D. (2002). "Equality or Priority." I: M. Clayton and A. Williams (Eds.). *The Ideal of Equality*. Basingstoke, Palgrave Macmillan: 81-125.

Det er også værd at bemærke, at et princip, som fordrer lighed i fordelingen af goder, er nært beslægtet med principper, som fordrer inklusion og diversitet. Eksempelvis kan et lighedsprincip, som fokuserer på fordeling af rettigheder og positioner, forstås som én variant af et princip, som foreskriver inklusion og diversitet.

Lighed i dataetikken vil typisk være relevant i forbindelse med at evaluere konsekvenserne af databehandling. Princippet vil tale imod databehandling, som fører til at relevante goder fordeles mindre lige. Princippet har således mere fokus på, hvordan databehandlingen samlet set påvirker personer, konkret på den fordeling af goder databehandlingen skaber, og mindre fokus på lighed i selve databehandlingen (se "Fairness" ovenfor).

Privatliv

Det engelske "privacy" kan på dansk bedst oversættes til "privatliv". Databehandling involverer per definition data, og disse data vil ofte være persondata, i nogle tilfælde også meget følsomme persondata. Privatliv i dataetik handler om at sætte etiske grænser for behandling af personlige data.[19]

For at kunne vurdere hvilken rolle privatliv spiller i dataetiske problemstillinger, må man behandle flere komplekse spørgsmål. Et første spørgsmål, som er genstand for omfattende debat i faglitteraturen, er, hvad det vil sige, at en person har privatliv? Et almindeligt bud er, at en person har privatliv med hensyn til personlige data, hvis andre ikke har eller har haft *adgang* til de pågældende data. Et andet bud er, at privatliv kræver, at en person har *kontrol* over adgang til de personlige data. Afhængigt af om privatliv forstås på den ene eller den anden måde, kan personers privatliv være udfordret i forskellige situationer.

[19] For et overblik over diskussionen af privatliv, se DeCew, J. (2018). "Privacy." I: E. N. Zalta (Ed.). *Stanford Encyclopedia of Philosophy*. For et overblik over diskussionen af privatliv specifikt i tilknytning til databehandling, se også Van Den Hoven, J., M. Blaauw, W. Pieters and M. Warnier (2020). "Privacy and Information Technology." I: E. N. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*.

Et andet spørgsmål er, hvilken etisk betydning vi skal tilskrive privatliv? Har privatliv eksempelvis intrinsisk eller instrumentel værdi? En plausibel teori er, at privatliv er instrumentelt værdifuldt, eksempelvis for at beskytte personer mod ydmygelse og afpresning og for at kunne skabe personlige relationer med varierende grader af intimitet.

Privatliv er også omdrejningspunktet i flere potentielle etiske principper, inklusive dataminimering, proportionalitet i databehandling og kontrol over personlige data. Dataminimering forstås typisk som et krav om, at agenten minimerer mængden af data, som databehandlingen behandler. Proportionalitet forstås som et krav om, at den gavnlige effekt af databehandlingen skal være tilstrækkeligt stor i forhold til mængden af data, som behandles. Endelig kan et princip om kontrol over data eksempelvis forstås som et krav om, at personer i udgangspunktet skal have en relevant mulighed for at afvise behandling af deres data.

Retfærdighed

Retfærdighed fremhæves i flere kilder som et centralt element i dataetikken. Retfærdighed i dataetik kan bedst fortolkes som, at databehandlingen respekterer de etiske krav, som de berørte personer kan stille *som* personer.

En fortolkning af retfærdighed er nødvendig, fordi retfærdighed er et komplekst og flertydigt begreb.[20] I nogle tilfælde bruges begrebet i så bred forstand, at det bliver synonymt med at handle sådan, som vi etisk set bør handle. I denne meget brede betydning er retfærdighed ikke et selvstændigt princip eller værdi, men en betegnelse for summen af de grunde som værdier og principper giver til at handle.

I en mere substantiel betydning drejer retfærdighed sig om netop det sæt af etiske principper, som handler om de etiske krav, personer kan stille til hinanden. Denne betydning er imidlertid fortsat en

[20] For et overblik, se Miller, D. (2017). "Justice." I: E. N. Zalta (Ed.). *Stanford Encyclopedia of Philosophy*.

formel kategori, som må gives mere konkret indhold – hvad er det helt præcist for etiske krav, som personer kan stille til hinanden som personer?

Når retfærdighed på denne måde gives mere konkret indhold, kan det vise sig at være identisk med et eller flere af de principper, som vi her har behandlet selvstændigt. Det kan eksempelvis vise sig, at et etisk krav, som personer kan stille til hinanden, er kravet om fairness eller respekt for autonomi.

Sikkerhed

Sikkerhed i dataetik handler om at sikre, at behandlingen af data ikke har utilsigtede og uønskede effekter.

Sikkerhed i databehandling kan betyde flere forskellige ting. Sikkerhed kan først og fremmest betyde sikkerhed i opbevaring og deling af data, således at data ikke utilsigtet bliver tilgængelige for uvedkommende personer. I denne forstand kan sikkerhed forstås som instrumentelt værdifuld, som redskab til beskyttelse af personers privatliv.

Sikkerhed kan også betyde, at databehandlingen er robust og pålidelig, således at kritiske systemer ikke bryder sammen eller begår fejl. I denne forstand kan sikkerhed være instrumentelt værdifuld for databehandleren som beskyttelse af systemets evne til at indfri databehandlerens mål med databehandlingen.

Endelig kan sikkerhed betyde, at databehandlingen garanterer overholdelse af visse krav, eksempelvis juridiske krav eller etiske grænser. I denne sammenhæng kan det for eksempel være, at databehandlingen anvender tekniske standarder for fairness, således at den påviseligt behandler personer lige i bestemte henseender.

Fordi sikkerhed og en række af de beslægtede begreber (for eksempel validering, robusthed, og pålidelighed) kan betyde væsentlig forskellige ting, er det afgørende at præcisere, hvad sikkerhed betyder i en konkret dataetisk problemstilling. Det er også værd at holde sig for øje, at sikkerhed

grundlæggende handler om at beskytte andre værdier eller principper. For at vurdere hvilken etisk betydning sikkerhed har, må man derfor præcisere de relevante værdier eller principper. Hvor vigtig sikkerhed etisk set er, afhænger derfor både af hvad man præcist mener med sikkerhed og af hvilken etisk vægt, vi bør give til at beskytte lige præcis det, som den pågældende form for sikkerhed handler om.

Velfærd

Velfærd er et udtryk for, hvor gode liv personer har. Velfærd i dataetik handler derfor om at holde sig for øje, hvordan databehandling påvirker personers liv og livskvalitet.

Velfærd forstås normalt som en intrinsisk etisk værdi.[21] Som sådan er velfærd centralt for en række etiske principper, som handler om personers liv og velbefindende, eksempelvis principper om godgørenhed og lighed.

Et første spørgsmål er, hvad vi mere præcist skal forstå ved, at en person har et godt eller dårligt liv. I faglitteraturen arbejder man overordnet med tre forskellige forståelser:

- Hedonisme hævder, at velfærd består i positivt ladede sindstilstande, for eksempel glæde, nydelse eller stolthed og fraværet af negativt ladede sindstilstande, for eksempel sorg, smerte eller skuffelse.
- Præference-teorien hævder, at velfærd består i opfyldelsen af personlige præferencer.
- Liste-teorien hævder, at velfærd består i opnåelsen af visse goder, som tilsammen udgør et godt liv, for eksempel selvindsigt, selvrespekt og meningsfulde, nære relationer.

Alle tre typer teorier kan formuleres i forskellige versioner, som varierer med hensyn til, hvordan de forstår detaljerne i teorien.

[21] For et godt overblik, se Crisp, R. (2017). "Well-being." E. N. Zalta (Ed.). *Stanford Encyclopedia of Philosophy*.

De tre typer teori kan i mange tilfælde nå til samme konklusion om, hvorvidt en handling skaber velfærd. Eksempelvis vil de fleste velfærdsteorier konkludere, at en handling, som påfører en person fysisk smerte, derved reducerer vedkommendes velfærd. Men der vil også være situationer, hvor forskellige teorier når til meget forskellige konklusioner om velfærd.

Et andet vigtigt spørgsmål er, hvordan man kan måle velfærd. Hvis velfærd er en central etisk værdi, så er det vigtigt, hvor *meget* velfærd personer har. Eksempelvis vil databehandling ofte kunne have som konsekvens, at nogle personers velfærd reduceres, mens andre personers velfærd øges, eller at en persons velfærd øges i én henseende, men reduceres i en anden. I sådanne situationer er det afgørende at kunne vurdere, om velfærden øges mere, end den reduceres, eller omvendt. Relevante teorier om, hvordan man kan kvantificere og måle velfærd, er for eksempel udviklet og anvendt i velfærdsøkonomi og medicinsk etik.

Endelig er det afgørende, hvor meget vægt hensynet til velfærd skal gives. Det vil sige, hvordan værdien af velfærd og de principper, som knytter sig til velfærd (for eksempel godgørenhed), skal afvejes mod andre værdier og principper, når disse trækker i forskellige retninger.

Værdighed

Værdighed forstås i etik som en egenskab, personer besidder, der fordrer at andre behandler dem med respekt. Værdighed i dataetik handler derfor om de dataetiske forpligtelser, som følger af respekt for personers etiske status som personer.[22]

For at kunne arbejde med værdighed i dataetik er man nødt til først at konkretisere, både hvad det vil sige at have værdighed, og hvilke etiske forpligtelser der følger heraf. Hvad vil det sige at respektere en persons værdighed, og hvilke handlinger viser ikke den fornødne respekt?

[22] For et godt overblik over forholdet mellem respekt og værdighed, se Dillon, R. S. (2018). "Respect." I: Zalta, Edward N. *Stanford Encyclopedia of Philosophy*.

Når det på denne måde er blevet gjort konkret, hvilken rolle værdighed antages at spille, må man efterfølgende vurdere, om det konkrete princip er plausibelt.

Konkrete principper om værdighed kan vise sig at trække på eller overlappe med andre begreber og principper. Eksempelvis kan man konkretisere princippet om respekt for værdighed ved at hævde, at denne respekt kræver, at agenter ikke krænker personers autonomi, eller at de behandler personer ligeligt i bestemte henseender (fairness).

Sammenfatning og afsluttende bemærkninger

I denne analyse har vi forsøgt at belyse tre centrale sider af dataetiske værdier og principper:

- Hvad etiske værdier og principper er,
- hvordan man kan identificere relevante værdier og principper i en konkret dataetisk problemstilling og
- hvilke etiske værdier og principper, der spiller en prominent rolle i den eksisterende litteratur om dataetik.

Analysen har vist, at dataetiske værdier og principper er komplekse begreber. Mange udtryk for værdier og principper optræder i flere forskellige betydninger i litteraturen, ligesom den samme værdi eller det samme princip kan blive henvist til og diskuteret med flere forskellige udtryk. Selv relativt entydige begreber er ofte teoretisk komplekse, således at det kræver detaljerede og vanskelige overvejelser at præcisere begrebet og slå fast hvilken etisk rolle, det spiller.

Analysen har identificeret i alt 13 værdier og principper, som spiller en væsentlig rolle i den eksisterende litteratur om dataetik. Præsentationen og diskussionen af disse illustrerer behovet for at præcisere, hvad man i en konkret kontekst forstår ved eksempelvis "autonomi", "fairness", eller "velfærd".

Analysen har argumenteret for, at der kan være fordele ved at foretage en sådan afklaring i forbindelse med behandlingen af konkrete dataetiske problemstillinger. Derved kan fokus begrænses til de dataetiske værdier og principper, som er relevante for den aktuelle problemstilling. En afklaring af værdier og principper i en konkret problemstilling identificerer de relevante værdier

og principper ved at analysere relevante begrundelser for synspunkter på problemstillingen og kritisk evaluere de værdier og principper, som optræder i sådanne begrundelser.

Samlet understreger de udfordringer, som analysen belyser, betydningen af det debatskabende og analytiske arbejde, som Dataetisk Råd har som opdrag i forhold til at kvalificere vores forståelse og diskussion af de dataetiske problemstillinger, vi som samfund kommer til at møde flere og flere af i de kommende år.

Kapitel 3

Dataetisk metode og teori

Dataetisk metode og teori

Dataetisk Råd arbejder med at analysere og skabe debat om dataetik. I disse år oplever Danmark en hastig vækst i behandlingen af data, både i den private og den offentlige sektor.[23] I nogle situationer kan de nye teknologiske muligheder rejse dataetiske problemstillinger. I disse situationer kan vi kun træffe gennemtænkte valg ved grundigt at overveje, hvad der er på spil. Dataetisk Råds analyser og debat af dataetiske problemstillinger kan således udgøre et vitalt bidrag til, at Danmark kan være på forkant med den teknologiske udvikling. Men hvordan analyserer og diskuterer man dataetik?

Dataetik handler om det sæt af problemstillinger, hvor personers behandling af data i sig selv udfordrer etiske værdier og principper (se kapitel 1 om "Hvad er dataetik?", kapitel 2 om "Dataetiske værdier og principper" og kapitel 4 om "Dataetiske problemstillinger"). For at kunne arbejde systematisk med analyse af dataetiske problemstillinger er det nødvendigt at anvende en metode tilpasset problemernes særlige etiske karakter.

Der findes ikke i faglitteraturen en veletableret metode til analyse af specifikt dataetiske problemstillinger. Der findes imidlertid en række etablerede metoder til analyse af etiske problemstillinger generelt, som er illustrereret ved at være anvendt på andre konkrete problemstillinger. Dataetisk Råd har fået udarbejdet denne analyse for at udvikle en metode til analyse af dataetiske problemstillinger, som redegør for og tilpasser de eksisterende metoder. Målet er at give Dataetisk Råd et udgangspunkt for reflekteret, systematisk arbejde med dataetiske problemstillinger.

Analysen kombinerer en fremstilling af dataetisk metode med en oversigt over prominente etiske teorier, som kan være relevante for dataetikken. Den skitserer indledningsvis anvendt etik som felt og skelner mellem henholdsvis en teoridrevet og en intuitionsdrevet tilgang til analysen af etiske problemstillinger. Analysen anbefaler en intuitionsdrevet tilgang og redegør i detaljer for de tre faser i en intuitionsdrevet analyse: 1) afdækning af problemstillingen, 2) kritisk vurdering af begrundelser

[23] I denne analyse anvender vi "behandling af data" som samlet betegnelse for de handlinger som på relevant vis vedrører data, f.eks. generation, indsamling, lagring, processering, analyse, anvendelse, og deling af data.

og 3) samlet afvejning af synspunkter. Afslutningsvis introducerer analysen syv prominente etiske teorier fra den moderne forskning, tre konsekvensetiske og fire deontologiske, som kan informere fortolkning og vurdering af begrundelser.

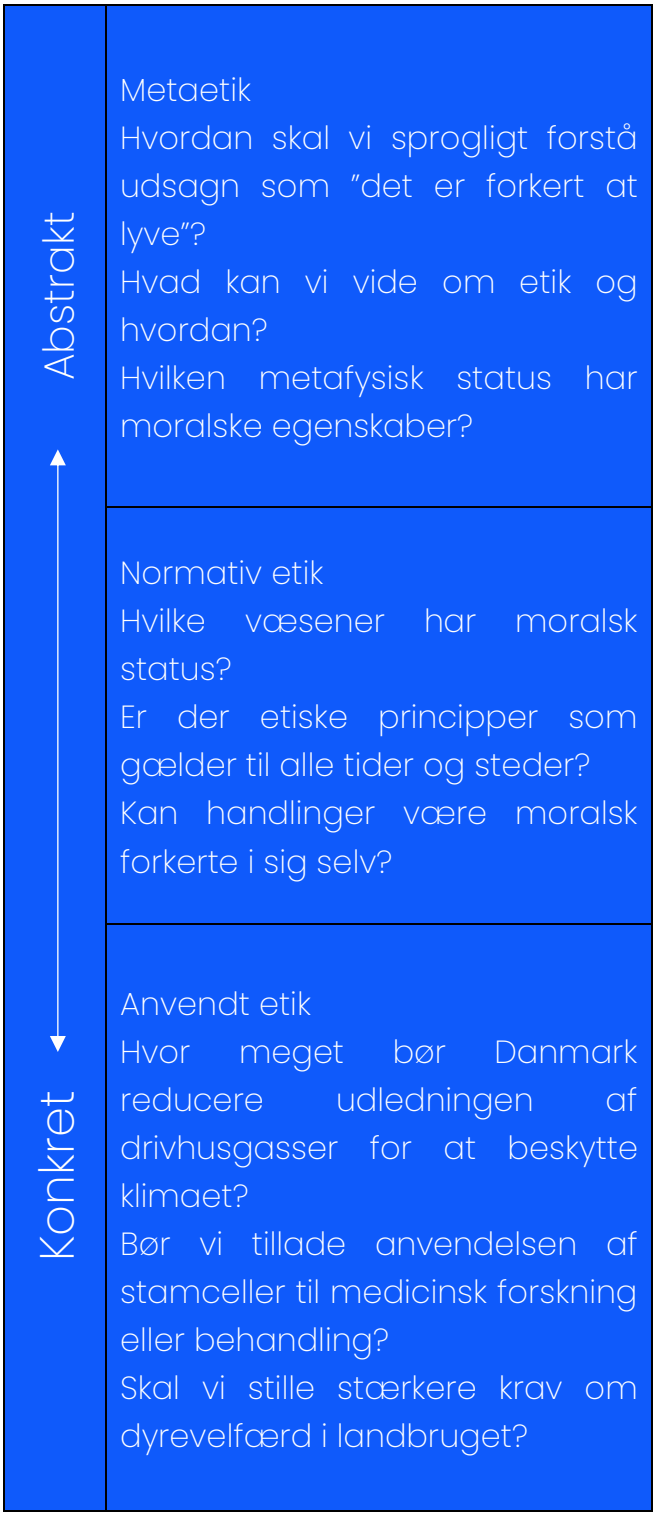
Dataetisk metode for Dataetisk Råd

Dataetisk Råds arbejde hører fagligt betragtet til genren "anvendt etik". I anvendt etik forsøger man at besvare konkrete spørgsmål om, hvordan agenter bør handle, typisk inden for et bestemt domæne for eksempel sundhedssystemet.

Derved adskiller den anvendte etik sig fra henholdsvis den normative etik, som forsøger at formulere helt generelle teorier om, hvordan man bør handle, og metaetikken, som forsøger at besvare spørgsmål om etik som videnskabeligt felt.

I den anvendte etik arbejder man med konkrete etiske problemstillinger. Inden for et domæne vil der ofte være en række handlinger eller praksisser, hvor handlingen eller praksissen udfordrer etiske værdier og principper (se kapitel 4 om "Dataetiske problemstillinger").

I almindelighed koncentrerer anvendt etik sig om at analysere de situationer, hvor det er uklart eller kontroversielt, hvordan man bør handle. I medicinsk etik forsøger man for eksempel at finde svar på spørgsmål som: bør vi lade sundhedspersonale udøve aktiv dødshjælp? Bør vi prioritere sundhedsbehandlinger afhængigt af behandlingens pris eller patientens alder? Bør vi tillade forældre at nægte deres børn livsvigtig behandling?



Dataetisk Råds arbejde kan tilsvarende fokusere på at analysere dataetiske dilemmaer: dataetiske problemstillinger hvor det er uklart eller kontroversielt, hvordan agenter bør behandle data.

Systematisk arbejde inden for den anvendte etik er typisk enten teoridrevet eller intuitionsdrevet, om end de to tilgange ofte kan supplere hinanden. Ved teoridrevet arbejde antager man en bestemt etisk teori som udgangspunkt og vurderer problemstillingen i lyset af denne teori. Ved intuitionsdrevet arbejde tager man udgangspunkt i begrundelser for og imod forskellige synspunkter på problemstillingen og vurderer i hvilken grad, de forskellige synspunkter kan støttes af gode begrundelser. Lidt forenklet kan man sige, at det første er en "top-down," og det andet er en "bottom-up"-tilgang til arbejdet med anvendt etik.

I det følgende skitseres først teoridrevet arbejde. Analysen påpeger i den forbindelse, at den teoridrevne tilgang til anvendt etik rejser visse udfordringer, som gør den mindre egnet til Dataetisk Råds arbejde. Derpå præsenteres intuitionsdrevet arbejde, inden analysen går i detaljer med processen og de centrale komponenter i denne tilgang til anvendt etik.

Teoridrevet arbejde med anvendt etik

Teoridrevet arbejde med anvendt etik betyder, at man tager udgangspunkt i en etableret generel teori om, hvordan agenter bør handle, og anvender den til at vurdere den konkrete problemstilling.

Inden for såkaldt "normativ etik" er der især siden midten af det 20. århundrede blevet udviklet en række sofistikerede generelle teorier om etik, for eksempel konsekvensetisk prioritarianisme, held-egalitarisme og kontraktualisme (se afsnittet neden for om dataetisk teori). Sådanne teorier præciserer hvilke forhold ved en situation, som påvirker, hvordan vi bør handle. Teoridrevet arbejde kan derfor tage udgangspunkt i en veludviklet teori, analysere, hvordan de

Eksempel: Dataetisk Råd overvejer hvordan en bestemt type samtykke til behandling af data bør udformes. Spørgsmålet analyseres i lyset af utilitarisme, som viser at samtykke spiller en rent instrumentel rolle, som en måde at begrænse visse former for etisk problematisk indsamling af data, samt i lyset af Kantianisme, som viser at bestemte former for samtykke spiller en væsentlig rolle for, hvordan data kan behandles. Begge teorier konkluderer, at det pågældende samtykke er utilstrækkeligt, og at databehandlingen er moralsk problematisk selvom personer samtykker.

relevante forhold optræder i netop den konkrete problemstilling, og konkludere, hvordan man bør handle.

Fordelen ved en teoridrevet tilgang til anvendt etik er således først og fremmest, at den kan trække på udførligt detaljeret teoretisk arbejde i faglitteraturen.

Ulemperne ved teoridrevet arbejde er, 1) at det forudsætter et detaljeret kendskab til den teori som skal anvendes, 2) at det begrænser konklusionens rækkevidde, fordi konklusionen kun gælder for så vidt, man antager den pågældende teori, samt 3) at det kan være teoretisk udfordrende at anvende generelle teorier på et konkret problem.

Den første ulempe følger direkte af fordelen ved teoridrevet arbejde. Anvendelsen af en detaljeret etisk teori forudsætter grundigt kendskab til den pågældende teori, og udvælgelsen af en relevant teori forudsætter kendskab til et udvalg af etiske teorier, samt den komplekse diskussion af deres respektive styrker og svagheder. Det kan således kræve en væsentlig investering af tid og ressourcer for personer, som ikke på forhånd har indgående kendskab til etiske teorier, at opnå forudsætningerne for at kunne påbegynde arbejdet.

Den anden ulempe følger af, at der i meget begrænset omfang findes konsensus i den normative etiske forskning. Det betyder, at der findes et stort antal konkurrerede etiske teorier, som hver for sig anerkendes af et begrænset antal forskere. I visse tilfælde vil den konklusion, som teoridrevet arbejde når til, være den samme konklusion, som man ville være nået til, hvis man havde antaget en eller flere andre etiske teorier. I sådanne tilfælde er der teoretisk konsensus om *konklusionen* på en etisk problemstilling. I mange tilfælde vil forskellige teorier imidlertid pege på forskellige konklusioner, og resultatet af en teoridrevet analyse vil derfor være af begrænset interesse for de, som ikke accepterer den pågældende etiske teori.

Den tredje ulempe følger af etiske teoriers karakter. Etiske teorier forsøger typisk at formulere principper for, hvordan vi bør handle, som er så abstrakte og generelle som muligt. Det betyder, at det i nogle tilfælde kræver vanskelig fortolkning at vurdere, hvordan en konkret problemstilling kan forstås i lyset af en etisk teori. Denne fortolkning kan kræve teoretiske forudsætninger fra normativ etik og metaetik og kan være kontroversiel, således at der ikke kan skabes enighed om, hvordan problemstillingen bør forstås i lyset af teorien.

Samlet medfører disse ulemper, at det i mange tilfælde kan være fordelagtigt for Dataetisk Råd at fokusere på en intuitionsdrevet tilgang til dataetiske problemstillinger.

Intuitionsdrevet arbejde med anvendt etik

Intuitionsdrevet arbejde med anvendt etik betyder, at man tager udgangspunkt i de synspunkter og begrundelser, som optræder i tilknytning til en konkret etisk problemstilling.

En intuition betyder i denne sammenhæng en umiddelbar fornemmelse af, at en etisk påstand eller et synspunkt på en etisk problemstilling er troværdigt. Tilgangen betegnes som intuitionsdrevet fordi, disse intuitioner spiller en vigtig rolle i den kritiske vurdering af begrundelser for synspunkter på problemstillingen.

Målet i den intuitionsdrevne tilgang til anvendt etik er kritisk at vurdere, hvilke synspunkter på en etisk problemstilling, som kan støttes af gode begrundelser. Det vil i mange tilfælde føre til, at man kan konkludere, at visse synspunkter ikke støttes af gode begrundelser, for eksempel fordi påstande i begrundelsen er utroværdige, eller fordi begrundelsen er et logisk ugyldigt argument. Samtidig kan der være andre synspunkter, som faktisk kan støttes af gode begrundelser. Det bedste svar på, hvordan agenter bør handle, antager den intuitionsdrevne tilgang må være det synspunkt, som samlet set bedst støttes af gode begrundelser.

Eksempel: Dataetisk Råd overvejer problemstillingen: "Hvornår kan data etisk forsvarligt bruges til at forudsige personers adfærd?" Spørgsmålet analyseres ved at samle, systematisere og analysere en række indvendinger mod brugen af data til at forudsige personers adfærd. Analysen viser at visse indvendinger, som umiddelbart kan virke overbevisende, hviler på tvivlsomme empiriske påstande eller ugyldige argumenter. De indvendinger som viser sig at være solide, peger på at det kan være etisk problematisk, at bruge bestemte typer data til at forudsige bestemte typer adfærd.

Fordelen ved en intuitionsdrevet tilgang er, 1) at den kan tage fat i en etisk problemstilling netop der, hvor eksempelvis den offentlige debat har engageret sig i problemstillingen, og 2) at arbejdet kræver færre teoretiske forudsætninger både at udføre og formidle.

Ulemperne ved intuitionsdrevet arbejde er, 1) at det ofte kræver et omfattende arbejde med at indsamle og systematisere relevante synspunkter og begrundelser, 2) at det kan være metodisk

udfordrende at foretage den kritiske vurdering af begrundelser og 3), at der kan være situationer, hvor flere modstridende synspunkter understøttes af nogenlunde lige gode begrundelser, og der derfor ikke er nogen entydig konklusion på, hvordan agenter bør handle.

Til trods for disse potentielle ulemper vil det i lyset af fordelene i mange situationer være fordelagtigt for Dataetisk Råd at arbejde intuitionsdrevet med en dataetisk problemstilling. I det næste afsnit diskuteres derfor processen med at foretage en intuitionsdrevet analyse af en dataetisk problemstilling. Analysen skitserer først den almindelige arbejdsproces og går derefter i detaljer med udvalgte dele af processen.

Hvordan arbejder man intuitionsdrevet med dataetik?

I lyset af de fordele og ulemper ved de to forskellige tilgange til anvendt etik, som er beskrevet ovenfor, forekommer det mest oplagt for Dataetisk Råd at fokusere på en intuitionsdrevet tilgang, suppleret med lejlighedsvis og fokuseret brug af teori.

Intuitionsdrevet arbejde med dataetik har til formål at afklare, hvilke synspunkter på en dataetisk problemstilling, som kan støttes af gode begrundelser, ved systematisk at afdække relevante synspunkter og begrundelser og kritisk vurdere begrundelsernes logiske gyldighed og troværdighed.[24]

I praksis kan processen med at foretage en intuitionsdrevet analyse af en dataetisk problemstilling inddeles i tre faser:

I den første fase afdækkes problemstillingen. Målet for den første fase er, at man klart og præcist kan formulere, hvad det er for en dataetisk problemstilling, som man ønsker at arbejde med, og har skabt sig overblik over relevante forhold ved problemstillingen, for eksempel hvordan databehandlingen rent teknisk fungerer. Samtidig bør man i denne fase redegøre for hvilke synspunkter på problemstillingen, som optræder i debatten, og hvordan de begrundes, og definere eventuelle yderligere synspunkter og begrundelser, som kan være relevante at vurdere.

I den anden fase underkastes de relevante begrundelser for synspunkter på problemstillingen en kritisk evaluering. Målet med den anden fase er at skelne mellem gode og dårlige begrundelser, herunder at kunne redegøre for *hvorfor* eventuelle dårlige begrundelser ikke er gode. I den forbindelse kan det være nødvendigt at afklare centrale begrebers betydning, at vurdere

[24] En god introduktion til metoden findes i Sommer Hansen, R. (2016). "Metode i normativ politisk teori." I: R. Sommer Hansen & S. Flinch Midtgaard (Eds.), *Metode i normativ politisk teori* (pp. 21-84). Viborg: Samfundslitteratur. Se også Ryberg, J. (2013). *Forstå etikken*. København: Hans Reitzels forlag.

argumenters logiske gyldighed samt at undersøge om de påstande, der indgår i begrundelserne, er troværdige.

I den tredje fase foretages en samlet vurdering af hvilke synspunkter, som støttes af gode begrundelser. Målet i denne fase er ideelt set at identificere et synspunkt, som i højere grad end andre synspunkter kan støttes med gode begrundelser. For at vurdere hvilke synspunkter, som bedst støttes at gode begrundelser, er det nødvendigt at veje forskellige begrundelser mod hinanden. Resultatet af denne afvejning er en konklusion på den dataetiske problemstilling, som viser hvilket synspunkt (eller synspunkter) på problemstillingen, der bedst støttes af gode begrundelser.

Denne fremstilling af processen er naturligvis idealiseret. I praktisk arbejde med en problemstilling kan de tre faser overlappe med hinanden, eller man kan opleve for eksempel, at man midt i arbejdet med at afveje synspunkter (fase 3) opdager et nyt relevant synspunkt, som man er nødt til at underkaste behandling jf. fase 1 og fase 2.

Som det også fremgår oven for, består hver af de tre faser af et eller flere metodiske skridt. I de følgende afsnit diskuteres nogle af de vigtigste metodiske skridt i den rækkefølge, hvor de optræder i processen.

Identifikation af relevante synspunkter og begrundelser

En dataetisk analyse har en dataetisk problemstilling som sin genstand. En dataetisk problemstilling er en bestemt form for databehandling, hvor databehandlingen i sig selv berører etiske værdier og principper. Det første væsentlige skridt i analysen, når man har identificeret den problemstilling man vil arbejde med (se kapitel 4 om "Dataetiske problemstillinger"), er at identificere relevante synspunkter på problemstillingen og relevante begrundelser for disse synspunkter. Et synspunkt på problemstillingen er en påstand om, hvordan databehandlingen bør foregå, herunder også for eksempel at man slet ikke bør udføre den pågældende form for databehandling. En begrundelse for et synspunkt er et argument med synspunktet som konklusion.

En almindelig måde at begynde afdækningen af synspunkter og begrundelser på er ved at se på, hvordan problemstillingen er blevet diskuteret både i faglitteraturen og i den offentlige debat (dagblade, radio, tv, etc.). Når problemstillingen er diskuteret, vil det typisk være for at forsvare et

bestemt synspunkt på problemstillingen, og i den forbindelse vil der normalt optræde en eller flere begrundelser for synspunktet.

I nogle tilfælde vil en problemstilling være beskedent behandlet, for eksempel fordi der er tale om et nyt fænomen. I sådanne tilfælde kan det være nødvendigt at forsøge at formulere plausible bud på, hvilke synspunkter personer *kunne* have på problemstillingen og hvilke begrundelser, som kunne fremføres til støtte for dem.

Endelig er den måske mest ambitiøse mulighed, at forsøge at afdække relevante synspunkter og begrundelser gennem selvstændige empiriske undersøgelser, for eksempel ekspertinterview med relevante forskere eller fokusgruppeinterview med repræsentative grupper af borgere.

Eksempel: Dataetisk Råd indsamler begrundelser for det synspunkt, at det er moralsk problematisk at anvende data til personaliserede og målrettede reklamer på sociale medier. To argumenter lyder henholdsvis, at denne anvendelse krænker personers autonomi, og at anvendelsen manipulerer med personers evne til at bestemme selv. De to begrundelser klassificeres som varianter af den samme indvending.

Det sidste skridt i kortlægningen af synspunkter og begrundelser er at systematisere de synspunkter og begrundelser, som man har identificeret. Meget ofte vil synspunkter og begrundelser, som optræder i forskellige kilder, kunne klassificeres som varianter af det samme synspunkt eller den samme begrundelse. Sådanne varianter kan med fordel behandles under et i den efterfølgende analyse.

Undersøgelse af relevante empiriske forhold

Et andet vigtigt skridt i den indledende afdækning af problemstillingen er at undersøge relevante empiriske forhold. Det kan for eksempel være, hvordan databehandlingen rent teknisk fungerer, hvilket omfang den har, og hvilke konsekvenser af databehandlingen man tidligere har observeret.

En afdækning af relevante empiriske forhold er vigtig, fordi mange begrundelser hviler på en eller flere påstande, som har empirisk indhold. Hvis de empiriske påstande, som findes eksplicit eller implicit i begrundelsen, er faktuelt forkerte, så svækker det på afgørende vis begrundelsen.

I nogle sådanne tilfælde vil man kunne ændre begrundelsen, således at den ikke længere hviler på en faktuelt forkert påstand. I disse tilfælde er etableret praksis at anvende "princippet om barmhjertighed". Dette princip foreskriver, at man altid forsøger at fortolke begrundelser således, at de bliver så gode som muligt. Princippet er motiveret af den indsigt, at det kun er ved at fokusere på den stærkest mulige fortolkning af begrundelser, at man sikrer at analysen ikke overser en afgørende begrundelse.

Eksempel: Dataetisk Råd analyserer en begrundelse som hævder, at en bestemt form for dataindsamling vil føre til en gradvis udhuling af privatlivsnormer, som på sigt vil gøre det lettere at udføre andre, moralsk problematiske typer dataindsamling. Rådet studerer internationale erfaringer med den pågældende form for dataindsamling, og konkluderer, at der ikke synes at være tegn på, at den har den påståede effekt på privatlivsnormer.

I praksis vil Dataetisk Råd kunne foretage vurderinger af centrale empiriske forhold ved oversigtsstudier af faglitteraturen og ved i nogle tilfælde at inddrage eksperter fra relevante fagmiljøer.

Afklaring af centrale begrebers betydning

I anden fase af processen er det overordnede mål at underkaste de relevante begrundelser et kritisk eftersyn for at skelne mellem gode og dårlige begrundelser. En almindelig oplevelse i denne del af processen er, at begrundelser trækker på centrale begreber, hvis betydning er uklar eller flertydig. I sådanne tilfælde er det vigtigt at afklare disse begrebers betydning for at kunne vurdere begrundelsen.

Udfordringen med uklare eller flertydige begreber i begrundelser for dataetiske synspunkter gør sig især gældende for begrundelser, der trækker på komplekse etiske begreber for eksempel "autonomi", "velfærd", eller "privatliv". I mange tilfælde vil det være uklart, hvordan den person, som har fremført begrundelsen, forstår begrebet, og derfor uklart hvordan man skal forstå begrundelsen. Der kan også være flere konventionelle betydninger af begrebet og derfor flere forskellige mulige varianter af begrundelsen, og nogle måder at forstå begrebet på kan føre til en bedre variant af begrundelsen end andre.

I de tilfælde, hvor det er uklart hvordan et begreb skal forstås, kan man forsøge at fastlægge begrebets betydning ved at trække på teoretiske forståelser, udviklet i den relevante faglitteratur, og ved at analysere begrebets betydning, som det forstås af relevante grupper af personer (se Kapitel 1 om "Hvad er dataetik?").[25]

I de tilfælde, hvor begrebet er flertydigt, er etableret praksis at skelne mellem de forskellige varianter af begrundelsen, som følger af disse forskellige betydninger. Efterfølgende bør man anvende princippet om barmhjertighed, således at man sikrer sig, at man behandler den fortolkning af begrebet, som fører til den bedst mulige variant af begrundelsen.

Eksempel: Dataetisk Råd analyserer en begrundelse, som hævder, at en bestemt form for databehandling krænker en moralsk ret til privatliv. Begrundelsen indeholder en etisk påstand om, at personer har en moralsk ret til privatliv, og en empirisk påstand om, at den pågældende form for databehandling reducerer personers privatliv. Ved at undersøge begrebet "privatliv" viser det sig, at den betydning af privatliv som optræder i den etiske påstand, ikke er den samme, som den betydning af privatliv der optræder i den empiriske påstand. Rådet konkluderer på den baggrund, at der ikke er tale om en god begrundelse.

Kritisk vurdering af begrundelsers logiske gyldighed

En god begrundelse for et synspunkt er et argument. Et argument består af to eller flere påstande (også kaldet "præmisser"), som tilsammen skal føre til argumentets konklusion. Konklusionen er i denne sammenhæng det synspunkt på en dataetisk problemstilling, som skal begrundes. For at påstandene kan føre til konklusionen, skal argumentet være logisk gyldigt. Et afgørende skridt i processen med at skelne mellem gode og dårlige begrundelser er derfor at vurdere, om argumentet faktisk er logisk gyldigt.

For at vurdere et argument skal man dels forstå argumentet præcist og dels vurdere, om konklusionen følger logisk af de påstande (præmisser), som argumentet rummer.[26] En præcis forståelse af argumentet kræver ofte, at man rekonstruerer og fremstiller det. En sådan

[25] Se også Gupta, A. (2019). "Definitions." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*, og Lippert-Rasmussen, K. (2016). "Begrebsanalyse i politisk teori." In R. Sommer Hansen & S. Flinch Midtgaard (Eds.), *Metode i normativ politisk teori* (pp. 163-184). Viborg: Samfundslitteratur.

[26] Metoder til analyse af argumenter er meget velbehandlede i faglitteraturen. En eksemplarisk introduktion er Fisher, A. (2004). *The Logic of Real Arguments*. Cambridge: Cambridge University Press.

rekonstruktion og fremstilling indebærer en kombination af omhyggelig læsning af de relevante kilder og konstruktiv fortolkning.

Rekonstruktion af argumenter er nødvendig fordi, det meget ofte forekommer, at argumenter er misvisende eller ukomplet formuleret. Det er eksempelvis almindeligt i den offentlige debat, at argumenter rummer en eller flere implicitte påstande. En implicit påstand er en nødvendig del af argumentet, som ikke optræder eksplicit, men underforstås, for eksempel fordi det antages, at påstanden er indlysende og ukontroversiel. Når et argument er afhængigt af sådanne implicitte påstande, så vil det typisk være hensigtsmæssigt at gøre dem eksplicitte i fremstilling af argumentet.

Et argument kan også være formuleret på en måde, som svækker begrundelsen, hvis formuleringen tages bogstaveligt. Når det i sådanne tilfælde alene er formuleringen, som svækker begrundelsen – det vil sige, at der findes en nærliggende omformulering af argumentet, som undgår de svagheder, den givne formulering medfører – så kan man med fordel antage, at formuleringen er misvisende. I disse tilfælde kan man rekonstruere argumentet ved at antage, at det tilsigtede argument er identisk med den nærliggende omformulering.

Rekonstruktionen af argumenter for at håndtere implicitte påstande og misvisende formuleringer er baseret på princippet om barmhjertighed i fortolkning. Når man skal vurdere, hvordan vi bør handle, er det relevant at kigge på de bedst mulige begrundelser for forskellige synspunkter. En rekonstruktion af et argument, som gør en implicit påstand eksplicit, tjener for eksempel til at sikre, at en potentielt god begrundelse ikke afvises alene fordi, en påstand var underforstået, da den blev formuleret.

Efter en eventuel rekonstruktion vil det ofte være en fordel at lave en præcis fremstilling argumentet. En sådan fremstilling skaber et overblik over argumentet, som kan støtte analysen af, om argumentet er logisk gyldigt. I faglitteraturen fremstilles argumenter ofte ved, at hver af argumentets påstande præsenteres for sig, og den logiske struktur mellem påstandene gøres eksplicit, men fremstillingen af et argument kan også have mere uformel karakter.

Når argumentet er rekonstrueret og fremstillet, er næste skridt at vurdere, om argumentet er logisk gyldigt. Et argument er logisk gyldigt, hvis konklusionen følger af påstandene (præmisserne). Det

betyder, at hvis påstandene er sande og argumentet gyldigt, så er konklusionen *nødvendigvis* også sand. Hvis argumentet omvendt er ugyldigt, så behøver konklusionen ikke at være sand, selvom påstandene er det, og argumentet kan derfor ikke begrunde synspunktet.

Hvis et argument er præcist fremstillet, så er det ofte enkelt at se, at argumentet er logisk gyldigt, når det er gyldigt. Det skyldes at logisk gyldige argumenter følger en af ganske få enkle strukturer, eksempelvis den såkaldte *modus ponens*:

1) Hvis A, så B
2) A

K) Derfor B

Et argument kan bestå af mere end to påstande og kan derfor have en mere kompleks samlet struktur. Men et gyldigt komplekst argument kan altid fremstilles som en serie af mere enkle delargumenter, der hver for sig har en af de enkle strukturer.

Eksempel: Dataetisk Råd vurderer om visse virksomheders indsamling af kundedata krænker personers privatliv. En virksomhed forsvarer praksissen med, at dataindsamlingen ville krænke privatlivet, hvis den gav anledning til udbredte protester fra kunderne, men at det ikke er tilfældet. Argumentet rekonstrueres som:

1) Hvis indsamlingen af data giver anledning til protester, så krænker den privatlivet.
2) Indsamlingen giver ikke anledning til protester.
3) Derfor krænker indsamlingen ikke privatlivet.

Analyse af argumentet viser at det er et eksempel på fejlslutningen "benægtelse af antecedenten". Argumentet er derfor ugyldigt.

Det er ofte vanskeligere at analysere argumenter, som er logisk *ugyldige*, og identificere, hvori problemet består. Som oven for nævnt, følger gyldige argumenter en enkel logisk struktur. Ugyldige argumenter kan afvige fra denne struktur på mange forskellige måder. Selv en analyse, som fokuserer på de mest almindelige fejl, må derfor forholde sig til en længere række fejlslutninger, blandt andet formelle slutningsfejl, for eksempel bekræftelse af antecedenten, cirkelslutninger, falsk ækvivalens, stråmand, autoritetsappeller, ad hominem og post hoc fejl.[27]

[27] Der er ikke i faglitteraturen konsensus om præcis hvordan man skal skelne mellem visse typer fejlslutninger. I praksis vil de fleste ugyldige argumenter imidlertid kunne fortolkes som varianter af ca. 5-10 meget almindelige fejlslutninger. Se Hansen, H. (2020). "Fallacies." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

Det er også vigtigt at holde sig for øje at den måde, man rekonstruerer et argument, kan påvirke, hvilke udfordringer argumentet møder. Det kan i mange tilfælde være uklart, om et ugyldigt argument fortolkes mest barmhjertigt på den ene eller den anden måde. I sådanne tilfælde kan fortolkningen påvirke hvilken type fejl, argumentet begår. Det er typisk også muligt at rekonstruere et ugyldigt argument på en måde, så argumentet bliver gyldigt ved at indsætte en manglende påstand eller omformulere en påstand. Normalt vil dette imidlertid føre til, at argumentet hviler på en eller flere påstande, som er indlysende utroværdige. Dermed må begrundelsen fortsat betegnes som dårlig.

Endelig er det værd at bemærke, at selvom analyse af argumenters logiske gyldighed kan formaliseres i nogen grad, så er der i vid udstrækning tale om et håndværk, som kræver træning.

Eksempel: Rådet bemærker i forbindelse med analysen af argumentet om, at dataindsamling ikke krænker privatlivet, fordi kunder ikke protesterer, at en mulig rekonstruktion af argumentet er:

1) Hvis indsamling af data ikke giver anledning til protester, så krænker den ikke privatlivet
2) Indsamlingen giver ikke anledning til protester.

K) Derfor krænker indsamlingen af data ikke privatlivet.

Rådets analyse viser, at argumentet i denne version er gyldigt, men den første påstand vurderes som utroværdig. Det virker ikke plausibelt, at indsamling af data *kun* krænker privatlivet, når den giver anledning til protester.

Kritisk vurdering af intuitioner om etiske påstande

En god begrundelse er et logisk gyldigt argument, hvor de enkelte påstande er troværdige. For at vurdere om en begrundelse er god eller dårlig må man derfor både vurdere, om argumentet er logisk gyldigt, og om de påstande, det indeholder, er troværdige. Troværdighed betyder i denne sammenhæng, at der er gode grunde til at acceptere påstanden.[28]

[28] Ofte kræver vi at påstande er *sande*. Dette kan imidlertid i nogle sammenhænge forekomme at være for stærkt et krav. Det kan således være vanskeligt at bevise, at selv meget troværdige empiriske påstande er sande, men forekomme urimeligt at afvise argumenter som hviler på sådanne påstande. Et krav om at påstande skal være sande ville også udfordre etiske påstande, fordi det er kontroversielt både om sådanne påstande overhovedet kan være sande, og hvordan man i givet fald kan bevise det.

En begrundelse for et synspunkt på en dataetisk problemstilling er et etisk argument, fordi konklusionen er et synspunkt om, hvordan vi bør handle. Et gyldigt etisk argument har den særlige karakter, at det er nødt til at indeholde en eller flere etiske påstande, som henviser til etiske værdier og principper. Det skyldes, at man ikke kan slutte alene fra empiriske påstande om, hvordan verden er, til en etisk konklusion om, hvordan vi bør handle. I arbejdet med at vurdere begrundelser for dataetiske synspunkter har man derfor brug for at kunne vurdere både empiriske og etiske påstande.

Det vil i mange tilfælde være enkelt at vurdere, om en empirisk påstand er troværdig. En empirisk påstand kan være umiddelbart troværdig, fordi den er alment anerkendt, for eksempel "det seneste årti er mængden af data som behandles vokset dramatisk". I de tilfælde, hvor en empirisk påstand ikke er umiddelbart troværdig, vil det være nødvendigt at vurdere påstanden i lyset af den relevante forskning. Eksempelvis vil det ofte være relevant i dataetikken at vurdere empiriske påstande om, hvordan teknologier til databehandling fungerer, og hvilke sociale og psykologiske effekter forskellige typer databehandling har. Undersøgelsen af sådanne relevante forhold er en essentiel del af afdækningen af den dataetiske problemstilling (se ovenfor).

Det er mere kompliceret at vurdere, om en etisk påstand er troværdig. En etisk påstand er en påstand, som centralt handler om etiske værdier og principper, for eksempel "det er etisk forkert at indsamle persondata uden samtykke". En sådan påstand kan ikke be- eller afkræftes empirisk. Én måde at gøre en etisk påstand troværdig er at argumentere for den. Et sådant argument vil imidlertid selv være et etisk argument med etiske påstande, hvis troværdighed skal vurderes. Belægget for en etisk påstand er derfor i sidste ende påstandens intuitive karakter.

En intuition skal, som tidligere nævnt, forstås som en umiddelbar opfattelse af, at en påstand eller et synspunkt er troværdigt.[29] Mange af de etiske værdier og principper, som er i centrum i faglitteraturen, er meget intuitive – de virker umiddelbart meget troværdige (se kapitel 2 om "Dataetiske værdier og principper"). De fleste vil for eksempel opleve en etisk påstand, som "vi

[29] For et overblik over den forskningsmæssige diskussion af intuitioner, se Pust, J. (2019). "Intuition." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

bør alt andet lige ikke handle på en måde, som påfører andre mennesker stærke lidelser", som umiddelbart troværdig.

Dét, at en etisk påstand er intuitiv, betyder imidlertid ikke, at vi uden videre skal acceptere den. Intuitioner kan og bør underkastes en kritisk vurdering. For at vurdere hvor pålidelig intuitionen er, bør man overveje tre væsentlige forhold:

1) Hvem og hvor mange deler intuitionen? I den udstrækning det kun er nogle personer, som opfatter påstanden som umiddelbart indlysende eller troværdig, mens andre ikke har en sådan opfattelse eller måske endda den modsatte opfattelse, så svækker det intuitionens pålidelighed.

2) Er der vildledende grunde til, at personer kunne opfatte påstanden som umiddelbart indlysende eller troværdig? Moralpsykologien har især de seneste årtier påvist en lang række irrelevante faktorer, som kan påvirke intuitioner om etiske påstande.[30] Hvis det er sandsynligt, at intuitionen er en reaktion på en sådan faktor, så svækker det intuitionens pålidelighed.

Eksempel: Dataetisk Råd analyserer en intuition knyttet til påstanden om, at deling af en bestemt type anonymiserede data om hygiejne er moralsk problematisk. Rådet overvejer generalisering af påstanden, og vurderer, at det ikke er umiddelbart indlysende eller troværdigt, at det i andre sammenhænge er moralsk problematisk at dele lignende data. En søgning i den moralpsykologiske faglitteratur sandsynliggør, at den oprindelige intuition er knyttet til forholdet omkring hygiejne, som et udslag af den såkaldte "yuck-factor". Rådet konkluderer på denne baggrund, at intuitionen ikke er pålidelig.

3) Kan intuitionen generaliseres til andre sammenhænge? Ofte viser det sig, at en etisk påstand, som forekommer umiddelbart troværdig i én sammenhæng, ikke forekommer troværdig i en anden sammenhæng. Når det er tilfældet, så svækker det intuitionens pålidelighed.[31]

Det kan være et omfattende arbejde at underkaste en intuition en grundig undersøgelse, som vurderer alle tre forhold. I praksis vil det derfor i arbejdet med en dataetisk problemstilling typisk

[30] Se eksempelvis Ryberg, J. (2016). "Eksperimentel filosofi, moralske intuitioner og metodologi." I: R. Sommer Hansen & S. Flinch Midtgaard (Eds.), *Metode i politisk teori* (pp. 217-228). Viborg: Samfundslitteratur.
[31] Se Holtug, N. (2016). "Værditeoriens metode." I: R. Sommer Hansen & S. Flinch Midtgaard (Eds.), *Metode i politisk teori* (pp. 199-216). Viborg: Samfundslitteratur.

være nødvendigt at udvælge enkelte intuitioner, som undersøges grundigt. Det kan være intuitioner, som Dataetisk Råd er i tvivl om, er pålidelige, eller intuitioner om etiske påstande, der spiller en særligt vigtig rolle i begrundelser for synspunkter på den dataetiske problemstilling.

Resultatet af en sådan undersøgelse vil være en vurdering af, hvor pålidelig intuitionen er. Det er i den forbindelse værd at bemærke, at modsat for eksempel argumenters logiske gyldighed, så er denne pålidelighed et gradsspørgsmål, snarere end et spørgsmål om at være eller ikke være pålidelig. I nogle tilfælde vil vi være tilbøjelige til at mene, at en intuition er så upålidelig, at vi ikke bør fæste lid til den. I andre tilfælde vil en intuition være delvist pålidelig, fordi der er visse forhold, som svækker den, men ikke nok til at vi helt vil afvise den. Af samme grund kan man ofte ikke skelne skarpt mellem gode og dårlige begrundelser. En begrundelse er bedre, jo mere pålidelige de relevante intuitioner er. Det betyder, at der kan være begrundelser for modstridende synspunkter på en dataetisk problemstilling, som alle må siges at være *gode nok* til, at vi bør tage dem med i vores overvejelser om, hvordan vi bør handle.

Samlet afvejning af begrundelser for synspunkter

Den tredje og sidste fase i arbejdet med en dataetisk problemstilling er en afvejning af begrundelser for og imod synspunkter på den dataetiske problemstilling. I de to tidligere faser har analysen afdækket problemstillingen og i den forbindelse identificeret relevante synspunkter og begrundelser, hvorpå disse begrundelser er blevet kritisk vurderet. Det betyder, at analysen i begyndelsen af denne fase har et overblik over både, hvilke synspunkter, man kunne have på den dataetiske problemstilling, og hvor gode begrundelser, der findes for hvert af disse synspunkter. Nu mangler kun den endegyldige vurdering af, hvordan dette stiller de forskellige synspunkter.[32]

Det første skridt er at eliminere de synspunkter, som ikke støttes af gode begrundelser. Analysen kan have medtaget synspunkter, for eksempel fordi de er blevet forsvaret i den offentlige debat eller fordi, de umiddelbart kan virke troværdige, som har denne karakter. Hvis en analyse behandler et bredt udsnit af synspunkter, så vil det ofte vise sig, at nogle synspunkter kun støttes af begrundelser,

[32] Et beslægtet skridt i processen diskuteres i faglitteraturen ofte under betegnelsen "bred reflekteret ligevægt". Se Daniels, N. (2016). "Reflective Equilibrium." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Processen med at etablere en sådan ligevægt er dog mere iterativ end den arbejdsproces som her er beskrevet, fordi den har til formål på samme tid at revidere både normativ etik (dvs. etisk teori) og vores vurderinger af konkrete etiske problemstillinger.

som i løbet af analysen er blevet afsløret som dårlige argumenter. Disse svagt begrundede eller ubegrundede synspunkter på den etiske problemstilling kan afvises.

Imidlertid vil det ofte være tilfældet, at flere konkurrerende synspunkter støttes af en eller flere gode begrundelser. I denne situation må man forsøge en samlet afvejning af, hvor stærke begrundelserne er for hvert af de konkurrerende synspunkter. På den baggrund kan man ideelt set konkludere, at et synspunkt er det bedst begrundede. I andre tilfælde vil man være tvunget til mere beskedent at konkludere, at der synes at være gode begrundelser for to eller flere synspunkter, og at det ikke er muligt på baggrund af den foreliggende analyse at konkludere, hvilket synspunkt som er at foretrække.

En afvejning af begrundelsers samlede styrke vil i praksis være et skøn, som blandt afhænger af intuitioner om både de enkelte begrundelser og deres samlede vægt. Selvom der i vid udstrækning er tale om et skøn, bør Dataetisk Råd i den forbindelse fastlægge procedurer for, at Rådsmedlemmerne individuelt og Rådet kollektivt kan foretage denne afvejning.

Etiske teorier for dataetikken

Ét vigtigt redskab i systematisk og reflekteret arbejde med dataetiske problemstillinger er kendskab til og anvendelse af relevant etisk teori. Selv i intuitionsdrevet arbejde, hvor analysen ikke tager udgangspunkt i en bestemt etisk teori, vil det ofte være relevant at inddrage etiske teorier. Teori kan for eksempel inddrages i vurderingen af begrundelser, hvor man kan støtte en påstands troværdighed ved at vise, at den følger af anerkendte etiske teorier. I dette afsnit introduceres nogle af de centrale teorier i moderne analytisk etik, som kan informere Dataetisk Råds arbejde.

Indledningsvis definerer afsnittet "etisk teori" og introducerer en skelnen mellem to familier af etiske teorier: konsekvensetiske teorier og deontologiske teorier. Disse to typer etiske teorier udgør de væsentligste teoretiske udgangspunkter for moderne analytisk anvendt etik.

Derpå præsenterer analysen en række indflydelsesrige og repræsentative etiske teorier fra hver af de to familier. Inden for konsekvensetikken præsenteres utilitarisme, egalitarisme og prioritarianisme. Inden for deontologien præsenteres held-egalitarisme, Kantianisme, dobbelt-effektdoktrinen og hypotetisk kontraktteori.

For hver af disse teorier påpeger analysen vigtige ligheder med og forskelle til andre teorier, ligesom den skitserer nogle af de argumenter, som ofte benyttes til at motivere eller kritisere teorierne.

Etik og etiske teorier

Etik drejer sig om, hvordan vi bør handle. Skal vi legalisere aktiv dødshjælp? Eller kriminalisere prostitution? Bør man holde ferie i Danmark i stedet for at flyve sydpå af hensyn til klimaet? Og skal man svare ærligt, når en kollega spørger, om den nye, kiksede frisure klæder hende? Sådanne etiske spørgsmål møder vi alle sammen. Etiske teorier er generelle svar på sådanne etiske spørgsmål. Det vil sige, at etiske teorier forsøger at forklare, hvad det er som i almindelighed og på tværs af forskellige situationer afgør, hvordan vi bør handle.

Den moderne, analytiske etik rummer en række forskellige etiske teorier. Især to familier af teorier er veludviklede og nyder bred opbakning i forskningen.

Den første familie af teorier lægger vægt på, hvilke *konsekvenser* en handling har. Sådanne teorier kaldes "konsekvensetiske":

- Konsekvensetik: Vi bør handle på den måde, som har de bedste konsekvenser.[33]

Den anden familie af teorier er vanskeligere at sammenfatte, men lægger typisk vægt på etiske *grænser,* som følger af eksempelvis hvilken *karakter,* handlingen har, eller hvilken *intention* den handlende har. Sådanne teorier kaldes "deontologiske":

- Deontologi: Vi bør handle på en måde, som overholder etiske grænser.[34]

Begge typer teori kan virke plausible. Det virker oplagt, at det spiller en rolle for, hvordan vi bør handle, om en handling har gode eller dårlige konsekvenser. Eksempelvis virker det som en god forklaring på, hvorfor vi ikke bør handle på en måde, som fører til menneskelig lidelse, at handlingen er forkert, *fordi* den har denne dårlige konsekvens. Tilsvarende er der tilfælde, hvor det kan virke oplagt, at vi ikke bør handle på en bestemt måde, fordi der er noget ved selve handlingen, som er problematisk. Eksempelvis oplever mange, at der er noget moralsk problematisk ved at skade en person, selv i situationer hvor det samlet set har gode konsekvenser, for eksempel fordi man derved forhindrer endnu mere skade på andre personer.

Konsekvensetik og deontologi er familier af etiske teorier. For at kunne sige noget konkret om hvordan vi bør handle, er man nødt til at anvende en mere præcis teori. I de følgende afsnit skitseres tre indflydelsesrige konsekvensetiske teorier og fire indflydelsesrige deontologiske teorier. Selvom de kun udgør et udsnit af teorierne i moderne analytisk etik, så kan de tjene som illustration af nogle af de vigtigste teoretiske forståelser og skillelinjer.

[33] For et overblik, se Holtug, N. (2014). "Konsekventialisme." I: A.-M. S. Christensen (Ed.), *Filosofisk etik* (pp. 77-97). Århus: Aarhus Universitetsforlag, og Sinnott-Armstrong, W. (2019). "Consequentialism." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

[34] Se Fogh Nielsen, C. (2014). "Deontologi." I: A.-M. S. Christensen (Ed.), *Filosofisk etik* (pp. 55-75). Århus: Aarhus Universitetsforlag, og Alexander, L., & Moore, M. (2020). "Deontological Ethics." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

Utilitarisme

Den mest berømte konsekvensetiske teori er utilitarisme. Utilitarismen hævder, at vi til enhver tid bør handle på den måde, som fører til den største sum af velfærd.[35] Utilitarisme er altså en etisk teori, som kombinerer et etisk princip med to dele af en værditeori:

- Konsekvensetik: vi bør handle på den måde, som har de bedste konsekvenser.
- Maksimering: de bedste konsekvenser er den situation, hvor *summen* af værdier er *størst*.
- Velfærd som værdi: den eneste etiske værdi er *velfærd*.

Det etiske princip, som udgør den første del af utilitarisme, er det generelle konsekvensetiske princip, at man bør handle på den måde, som har de bedste konsekvenser. Fordi konsekvensetiske teorier handler om gode og dårlige konsekvenser, er det afgørende for disse teorier, at de har en værditeori (aksiologi).

Den anden del af utilitarisme er således den værditeoretiske ide om maksimering. Maksimering betyder, at konsekvenser skal vurderes efter, hvor stor summen af værdier er. Hvis man skal vælge mellem to handlinger, så skal man altså forestille sig, hvordan verden vil se ud, hvis man handler på henholdsvis den ene og den anden måde. For hver af de to måder verden kan komme til at se ud, skal man lægge alle etiske værdier sammen. De bedste konsekvenser er ifølge maksimering den situation, hvor summen af værdier er størst.

Den tredje del af utilitarismen er den værditeoretiske ide, at velfærd og *kun* velfærd har etisk værdi. Det betyder, at vi skal vurdere konsekvenser, ved at kigge på hvor meget velfærd de skaber. Velfærd betyder generelt, om et liv går godt eller dårligt, men der findes forskellige måder at præcisere hvad velfærd betyder (se Dataetisk Råds analyse "Dataetiske værdier og principper"). Der findes derfor forskellige varianter af utilitarisme, som benytter forskellige velfærdsteorier.

Fortalere for utilitarisme fremfører at det er overordentlig plausibelt at konsekvenser spiller en vigtig rolle for hvordan vi bør handle, samt at velfærd er en etisk værdi. Teorien har endvidere den fordel, at det i princippet er muligt at give et enkelt og præcist svar på hvordan man bør handle i en hvilken

[35] Utilitarismen har rødder i den Britiske oplysningstid, hvor den især blev udviklet af filosofferne Jeremy Bentham, John Stuart Mill og Henry Sidgwick. For et overblik, se: Driver, J. (2014). "The History of Utilitarianism." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

som helst situation. Utilitarisme har haft stor indflydelse, for eksempel i velfærdsøkonomi og medicinsk etik.

Kritikere vil på den anden side bemærke, at utilitarisme er en krævende etisk teori. Den medfører, at der ofte kun er netop én måde man bør handle på, fordi de fleste mulige handlinger har i hvert fald lidt forskellige konsekvenser. Der er kun flere forskellige handlinger at vælge mellem i netop de situationer, hvor mindst to handlinger har præcis lige gode konsekvenser, samtidig med at disse konsekvenser er bedre end alle andre handlingers konsekvenser.

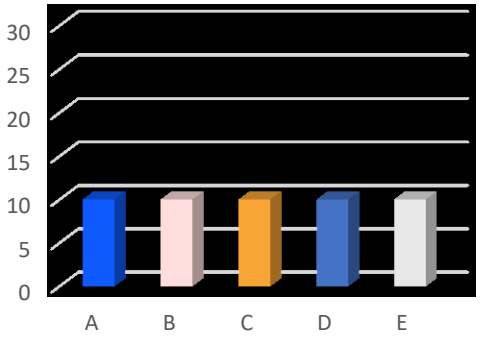
Utilitarisme er også krævende fordi den er upartisk. Alles velfærd tæller præcis lige meget, og man bør handle på den måde som har de bedste konsekvenser. Det kan for eksempel betyde, at man bør handle på en måde, som har meget dårlige konsekvenser for én selv, fordi handlingen har gode konsekvenser for andre.

Endelig giver utilitarisme i nogle situationer svar på hvordan vi bør handle, som er kontroversielle. De klassiske eksempler omhandler situationer, hvor en agent er tvunget til at vælge mellem at gøre skade på én person, for at forhindre skade på en gruppe andre personer, eller lade de sidstnævnte komme til skade. Hvis man minimerer den samlede skade ved at skade den første person, så bør man ifølge utilitarismen ofre vedkommende.

Egalitarisme

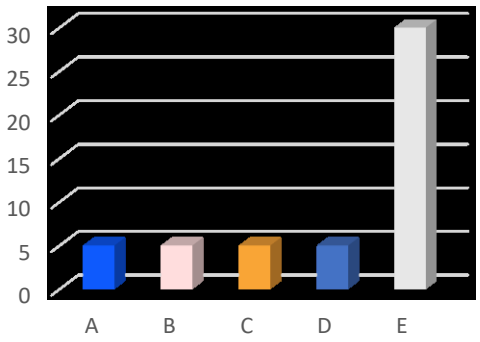
Vi har ovenfor set, at utilitarismen hævder at de bedste konsekvenser er den situation, hvor summen af værdier er størst (maksimering). Det betyder, at det ifølge utilitarismen er lige meget, hvordan værdierne er fordelt – det eneste der tæller er hvor meget velfærd der findes.

1



I mange situationer kan det imidlertid virke som om, at det spiller en vigtig rolle, om værdier er ligeligt eller ulige fordelt. Se eksempelvis de to figurer til højre (situation 1 og 2). Vi forestiller os, at A-E er personer, og at søjlerne repræsenterer de værdier hver person har, for eksempel mængden af velfærd. Hvordan bør vi handle, hvis vi skal vælge mellem to handlinger, som fører til henholdsvis situation 1 og situation 2?

2



Ifølge utilitarisme er det lige meget hvordan vi handler. Summen af værdier er i begge situationer 50, og der er derfor ingen grund til at foretrække den ene situation frem for den anden situation. Det kan virke urimeligt – bør vi ikke handle på den måde, som fører til den lige fordeling (1), snarere end på den måde som fører til den meget ulige fordeling (2)?

En prominent etisk teori, som tilgodeser denne intuition, er konsekvensetisk egalitarisme. En stærk variant af konsekvensetisk egalitarisme hævder at vi bør handle på den måde, som skaber den mest lige fordeling af goder mellem personer.[36] Konsekvensetisk egalitarisme kombinerer således to etiske principper:

- Konsekvensetik: vi bør handle på den måde, som har de bedste konsekvenser.
- Lighed: de bedste konsekvenser er den situation, hvor *fordelingen* af værdier er *mest lige*.

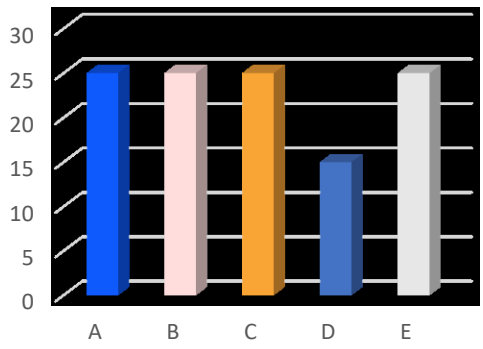
[36] Moderne egalitarisme forsvares blandt andet af den amerikanske filosof Larry Temkin. For et overblik over egalitarisme, se: Arneson, R. J. (2013). "Egalitarianism." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*, og Flinch Midtgaard, S. (2016). "Lighedsteorier og metode." I: R. Sommer Hansen & S. Flinch Midtgaard (Eds.), *Metode i politisk teori* (pp. 101-111). Viborg: Samfundslitteratur.

Ligesom utilitarisme kan egalitarisme kun vurdere konkrete situationer, hvis den kombineres med en værditeori (aksiologi). Det kan være den teori, at kun velfærd har værdi, som utilitarismen benytter, men det behøver det ikke. Egalitarisme kan også kombineres med andre værdier end velfærd, eller en værditeori som hævder, at der findes flere forskellige moralske værdier.

Prioritarisme

Konsekvensetisk egalitarisme er ligesom utilitarisme en krævende teori. Den siger at vi altid bør handle på den måde, som fører til den mest lige fordeling af værdier. Mange egalitarister er derfor fortalere for en svagere version af lighedsprincippet. Det kan for eksempel sige, at konsekvenserne af en handling i *én henseende* er bedre, når fordelingen af værdier er mere lige.

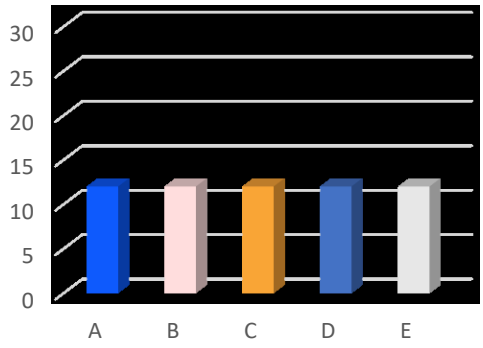
3



Selv en svækket udgave af egalitarisme kan imidlertid virke for krævende. Egalitarisme er sårbar over for den den såkaldte "levelling down"-indvending:

- "Levelling-down"-indvendingen: Det er ikke på nogen måde bedre, at skabe mere lighed, når *ingen* stilles bedre derved.

4



Se eksempelvis de to figurer ovenfor (situation 3 og 4). Vi forestiller os ligesom i forrige eksempel, at A-E er personer, og at søjlerne viser fordelingen af værdier, for eksempel velfærd. I situation 3 har de fleste det altså relativt godt, men én person har det værre end de andre. I situation 4 har alle det lige godt, men alle er dårligere stillet end i 3. Spørgsmålet er så: hvis vi kan ændre situation 3 til situation 4, bør vi så gøre det?

Egalitarister er nødt til at sige, at der er i hvert fald én grund til at foretrække situation 4, nemlig at den er mere lige. Kritikere af egalitarisme påpeger, at det virker absurd. Når selv de dårligst stillede (D) ikke får det bedre, eller endda får det værre, så er det svært at se, at der skulle være nogen som helst grund, til at ændre situation 3 til situation 4.

Èn teori som undgår denne udfordring, men fortsat giver mulighed for at tage hensyn til fordelingen af værdier, er priotarianisme.

Prioritarianisme hævder at konsekvenser skal vurderes på den måde, at værdier har faldende etisk betydning, jo mere af værdien en person har.[37] Det er altså etisk set vigtigere, at give mere værdi til en person, som ikke har meget af værdien, end til en person, som allerede har meget. I modsætning til egalitarisme, så skyldes dette imidlertid ikke at der derved kommer mere lighed. Etiske værdiers vægt er uafhængig af hvordan andre personer er stillet – den afhænger alene af hvordan den enkelte person er stillet. Konsekvensetisk prioritarianisme kan derfor forstås som en kombination af tre principper:

- Konsekvensetik: vi bør handle på den måde, som har de bedste konsekvenser.
- Vægtet maksimering: de bedste konsekvenser er den situation, hvor summen af *vægtede* værdier er størst.
- Faldende marginalvægt: værdier vejer tungere jo mindre af værdien en person har, og mindre tungt jo mere af værdien en person har.

Ligesom konsekvensetisk egalitarisme kræver prioritarianisme en værditeori, og ligesom egalitarisme kan den både antage at velfærd er den eneste værdi, eller antage en mere kompleks værditeori.

Held-egalitarisme

I det ovenstående har analysen skitseret tre konsekvensetiske teorier. I de følgende afsnit introduceres teorier fra den anden store familie – deontologien. Vi har ovenfor skitseret konsekvensetisk egalitarisme. En vigtig deontologisk variant af egalitarismen er såkaldt held-egalitarisme. Held-egalitarisme forsøger at tage højde for, at nogle former for ulighed kan være selvforskyldte. Det er intuitivt mindre moralsk problematisk, eller måske slet ikke problematisk, at værdier er ulige fordelt, hvis de personer som har det dårligere selv er ansvarlige for, at de har det dårligere. Held-egalitarisme formuleres ofte som følgende princip:

[37] Prioritarianisme er især udviklet af den britiske filosof Derek Parfit. Se diskussionen af prioritet og prioritarianisme i Arneson, R. J. (2013). "Egalitarianism." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*, og Holtug, N. (2016). "Værditeoriens metode." I: R. Sommer Hansen & S. Flinch Midtgaard (Eds.), *Metode i politisk teori* (pp. 199-216). Viborg: Samfundslitteratur.

- Held-egalitarisme: Det er uretfærdigt, hvis nogle personer er stillet dårligere end andre, uden at være ansvarlige for at være stillet dårligere.[38]

En klassisk illustration af den intuition som kan ligge til grund for teorien er Æsops fabel om myrerne og græshoppen. I fablen ender græshoppen med at have det dårligere end myrerne, fordi den har brugt hele sommeren på at nyde livet, mens myrerne har slidt for at samle mad og bygge hi til den kolde vinter. I sådanne situationer, hævder held-egalitarisme, er uligheden ikke uretfærdig, og der er derfor ikke en etisk grund til at skabe lighed.[39]

Omvendt fremhæver fortalere for held-egalitarisme, at teorien kan forklare hvorfor mange af de uligheder, som vi typisk opfatter som moralsk problematiske, er uretfærdige. Eksempelvis kan teorien forklare at sexisme er moralsk problematisk ved at påpege, at et samfund, hvor kvinder systematisk stilles dårligere end mænd *fordi* de er kvinder, skaber uligheder som kvinder ikke er ansvarlige for, og at dette er uretfærdigt. Endelig er det værd at bemærke, at held-egalitarisme typisk ikke formuleres som en konsekvensetisk teori. Teorien hævder ikke, at vi bør handle på netop den måde, som har de bedste konsekvenser, men kun at der findes en bestemt type ulighed, som vi har meget stærke grunde til at forhindre. Held-egalitarisme kan derfor bedre forstås som et deontologisk princip, som sætter visse etiske grænser for hvordan vi bør handle.

Kantianisme

Den nok mest indflydelsesrige deontologiske teori er Kantianisme.[40] Let parafraseret er de to mest indflydelsesrige formuleringer:

- Det kategoriske imperativ 1: vi bør handle således, at vi rationelt kan ville, at maksimen for vores handling blev en universel lov.

[38] Held-egalitarisme udvikles i sidste del af det 20. århundrede af blandt andre filosofferne John Rawls, Ronald Dworkin og Gerald Cohen. For et overblik se: Lippert-Rasmussen, K. (2018). "Justice and Bad Luck." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

[39] Bemærk at held-egalitarisme er forenelig med, at der kan være andre etiske grunde til at hjælpe. Det kan eksempelvis være at myrerne bør hjælpe græshoppen af barmhjertighed eller godgørenhed. Pointen er blot at der ikke er en grund til at hjælpe, som handler om at skabe *lighed*.

[40] Kantianisme er navngivet efter sin ophavsmand, den tyske oplysningsfilosof Immanuel Kant, men er siden fortolket og udviklet på mange forskellige måder. Eksempelvis trækker nogle fortalere for held-egalitarisme, som vi ovenfor har præsenteret, eksplicit på Kant's moralfilosofi. For et overblik, se: Johnson, R., & Cureton, A. (2019). "Kant's Moral Philosophy." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

- Det kategoriske imperativ 2: vi bør handle således, at vi behandler andre personer som mål i sig selv, ikke blot som middel for vores egne mål.[41]

Kant hævdede selv at der er tale om forskellige formuleringer af det samme princip. Mange forskere har imidlertid fortolket formuleringerne som substantielt forskellige etiske principper.

Det kategoriske imperativ 1 kræver, at den maksime, som ligger til grund for en handling, rationelt kan gøres til universel lov. En maksime forstås i den forbindelse som et princip for at handle, og en universel lov som en lov der binder alle personer. Kant hævder, at eksempelvis løgn og tyveri forbydes af princippet, fordi det ville underminere maksimen for løgn eller tyveri at gøre den til universel lov. Muligheden for at lyve forudsætter eksempelvis, at vi har tillid til hinanden, fordi vi taler sandt det meste af tiden, og tyveri forudsætter tilsvarende, at der findes privat ejendom. Hvis løgn og tyveri blev til universel lov, således at alle forsøgte at lyve og stjæle konstant, så ville forudsætningerne for løgn og tyveri forsvinde. Derfor kan man ifølge Kantianisme ikke rationelt ville, at maksimen for en løgn eller et tyveri blev til universel lov.

Det kategoriske imperativ 2 kræver, at man i sine handlinger behandler andre personer som mål i sig selv, ikke blot som middel for sine egne mål. Princippet er baseret på det synspunkt, at autonome væsener fortjener et særligt hensyn til deres status som autonome væsener. Hvis man ignorerer dét, at andre personer er væsener med viljen til at sætte sig mål for deres liv, og behandler dem som ting der alene er til for at tjene sine egne mål, så overskrider man en etisk grænse. Det er vigtigt i den forbindelse at holde sig for øje, at Kantianisme ikke forbyder at bruge personer som middel til sine mål. Det gør mennesker hele tiden i almindelige sociale interaktioner. Princippet forbyder alene at bruge personer *kun* som middel.

Kantianisme har været genstand for en enormt omfattende debat i faglitteraturen. Der er således uenighed både om hvordan teorien skal fortolkes, og om hvorvidt den er plausibel. Fortalere vil

[41] Den korrekte fortolkning af Kant's teori er genstand for meget omfattende debat i forskningen, blandt andet fordi Kant formulerer hvert princip flere gange, på lidt forskellige måder. To af Kants formuleringer af henholdsvis den første og den anden version af det kategoriske imperativ lyder: „handle so, daß die Maxime deines Willens jederzeit zugleich als Prinzip einer allgemeinen Gesetzgebung gelten könne" (*Kritik der praktischen Vernunft* (1788)) og „handle so, dass du die Menschheit sowohl in deiner Person, als in der Person eines jeden anderen jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst" (*Grundlegung zur Metaphysik der Sitten* (1785)).

fremhæve, at teorien giver en forklaring på vores moralske forpligtelser, som tager udgangspunkt i en sofistikeret analyse af vores natur som moralske væsener.

Kritikere har især påpeget, at det i forbindelse med den første version af det kategoriske imperativ kan være vanskeligt at definere den relevante maksime for en handling, samt at kravet i den anden formulering, om at man skal behandle andre personer som mål i sig selv, risikerer at være et indholdsløst, formelt krav om, at tage de hensyn, som vi etisk skylder andre, uden at disse hensyn bestemmes af teorien.

Hypotetisk kontraktteori

Kantianisme er, som ovenfor nævnt, blevet fortolket på mange måder, og har inspireret en lang række deontologiske teorier. Én type moderne deontologisk teori, som trækker på den første formulering af det kategoriske imperativ, er hypotetisk kontraktteori.

Hypotetiske kontraktteorier tager udgangspunkt i den ide, at vi bør handle på måder som alle rationelt kunne blive enige om bør være tilladt.[42] Teorierne er hypotetiske, fordi der ikke er tale om, at personer *faktisk* indgår aftaler med hinanden om hvordan vi bør handle, men om hvilke aftaler det ville være *muligt* at indgå. Teorien fokuserer også på hvad personer *rationelt* ville aftale, ikke på hvad personer i praksis ville acceptere. I virkeligheden kan personer være for snæversynede, kortsigtede eller selvoptagne til at blive enige om de principper, som de rationelt kunne aftale. En indflydelsesrig version siger:

- Kontraktualisme: vi bør handle således, at ingen rimeligvis kunne protestere imod vores handling.[43]

Hvordan ved man, om nogen rimeligvis (eng. "reasonably") kunne protestere imod en handling? Kontraktualisme antager, at en rimelig protest opfylder to krav. For det første skal protesten kunne begrundes, hvilket betyder at den kan henvise til et etisk princip, som handlingen strider imod. For det andet skal det princip, som protesten kan henvise til, ikke selv kunne rimeligt afvises. Det betyder

[42] For et overblik, se: Ashford, E., & Mulgan, T. (2018). "Contractualism." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.
[43] Denne version er udviklet af den amerikanske filosof Thomas Scanlon.

ikke, at alle personer rationelt *skal* acceptere princippet. Men det betyder at der ikke må være nogen, som rimeligt kan sige nej til princippet, for eksempel fordi det ikke tager hensyn til deres interesser.

Fortalere for kontraktualisme fremhæver, at teorien er mindre krævende end konsekvensetiske teorier, men ligesom konsekvensetikken kan forklare hvorfor konsekvenser spiller en vigtig rolle for hvordan vi bør handle.

Kritikere har til gengæld fremført, at teorien forudsætter en yderligere forklaring af hvad det vil sige, at nogen rimeligvis kan afvise et etisk princip, og at det kan være svært at give sådan en forklaring uden en selvstændig teori om hvordan vi bør handle, som ville gøre kontraktualisme overflødig.

Doktrinen om dobbelteffekt

En anden indflydelsesrig deontologisk teori som er inspireret af Kantianisme trækker på den anden version af det kategoriske imperativ. Den såkaldte doktrin om dobbelteffekt (DDE) hævder, at en handling er *særligt* etisk forkert, hvis den handlende har til intention at forårsage en etisk dårlig konsekvens, for eksempel at skade en person:

- Dobbelteffektdoktrinen: Vi bør ikke intentionelt forårsage noget etisk dårligt, enten ved at denne konsekvens er et mål i sig selv for vores handling, eller ved at konsekvensen er et middel til at opnå et mål for vores handling.[44]

DDE får sit navn af, at princippet implicit hævder, at der er en etisk forskel mellem to forskellige måder at forårsage en etisk dårlig konsekvens: intentionelt og blot forudset. Dårlige konsekvenser som kun er en såkaldt "forudset side-effekt" fordømmes ikke af princippet. I et klassisk eksempel giver en læge smertestillende behandling til en terminal patient med ubærlige lidelser. Lægen intenderer at lindre patientens smerter, men forudser også, at patienten derved dør. I sådanne tilfælde har handlingen ifølge teorien en "dobbelteffekt": en intenderet god effekt (smertelindring) og en uintenderet men forudset dårlig effekt (patienten dør). Fordi den dårlige konsekvens (patienten dør) kun er forudset forbyder princippet ikke, at lægen udfører handlingen. Omvendt forbyder princippet dødshjælp, hvor en læge slår en terminal patient ihjel *for at* forhindre patientens lidelser.

[44] For et overblik, se: McIntyre, A. (2019). "Doctrine of Double Effect." I: E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

Det er vigtigt at holde sig for øje, at princippet ikke benægter, at uintenderede dårlige konsekvenser taler imod en handling. I eksemplet med lægen som giver smertelindring er patientens uintenderede død naturligvis en dårlig konsekvens, som taler imod at lægen bør give smertelindring. Til gengæld hævder tilhængere af DDE, at grundene til at undgå uintenderet skade ikke vejer nær så tungt som grundene til at undgå intenderet skade. Det er denne forskel som forklarer, at det i nogle tilfælde kan være legitimt at forårsage uintenderet skade. I eksemplet med lægen er det fordi patientens død kun er forudset, at den gode konsekvens (smertelindring) kan opveje den dårlige, således at lægens handling samlet set bliver etisk forsvarlig.

Fortalere for DDE fremfører, at princippet synes at give intuitive svar i mange situationer, hvor konsekvensetiske teorier hævder, at vi bør gøre intenderet skade. DDE vil eksempelvis, i modsætning til almindelige konsekvensetiske teorier, fordømme at vi intenderet dræber én uskyldig person for at redde to andre.

Kritikere har påpeget, at det kan være vanskeligt at gøre forskellen på intenderede og blot forudsete konsekvenser præcis, samt at DDE også synes at give meget kontraintuitive svar i nogle situationer, eksempelvis i en situation hvor vi kan redde to personer ved at ændre én persons død fra blot forudset til intenderet (men ikke forhindre at vedkommende dør).

Kapitel 4

Dataetiske problemstillinger

Dataetiske problemstillinger

I disse år oplever Danmark en hastig vækst i behandlingen af data, både i den private og den offentlige sektor.[45] Udviklingen drives af en kombination af stadigt lettere adgang til stadigt større mængder data og stadigt mere sofistikerede værktøjer til at analysere data, som tilsammen udvider mulighederne for at bruge data til at løse meget forskellige udfordringer i mange sektorer.

I nogle situationer kan de nye muligheder for databehandling rejse dataetiske problemstillinger. I disse situationer kan vi kun træffe gennemtænkte valg om hvordan vi bør udnytte mulighederne, ved grundigt at overveje, hvad der etisk er på spil. Dataetisk Råd arbejder med at analysere og skabe debat om dataetik. Rådets analyser og debat af dataetiske problemstillinger kan således udgøre et vigtigt bidrag til, at Danmark kan være på forkant med den teknologiske udvikling. Men hvordan identificerer og analyserer man en dataetisk problemstilling?

Dataetik handler om det sæt af problemstillinger, hvor personers behandling af data i sig selv udfordrer etiske værdier og principper (se Dataetisk Råds analyse "Hvad er dataetik?"). Der findes imidlertid ikke en etableret forståelse af, hvilke problemstillinger dataetikken omfatter. Det skyldes dels at teknologien udvikler sig, og rejser nye udfordringer i takt med at der åbner sig nye muligheder for databehandling, og dels at det først er i de seneste år, at dataetik er blevet genstand for fokuseret opmærksomhed i forskningen og den offentlige debat.[46]

For at identificere en dataetisk problemstilling er man derfor nødt til at vurdere, om en given form for databehandling udfordrer etiske værdier og principper (se Dataetisk Råds analyse "Dataetiske værdier og principper"). Når en problemstilling er identificeret, må man kritisk evaluere argumenter for og imod forskellige mulige syn på problemstillingen (se Dataetisk Råds analyse "Dataetisk teori og metode"). Dataetisk Råd har fået udarbejdet denne analyse for at præsentere centrale

[45] I denne analyse anvender vi "behandling af data" som samlet betegnelse for de handlinger som på relevant vis vedrører data, f.eks. generation, indsamling, lagring, processering, analyse, anvendelse, og deling af data.
[46] Det er dog værd at bemærke, at en række etiske problemstillinger som knytter sig til databehandling, er blevet mere omfattende behandlet. Det gælder eksempelvis overvågnings- og privatlivsproblemer, samt visse problemer som knytter sig til kunstig intelligens. I de tilfælde, hvor der findes en sådan forudgående forskning, bør dataetikken naturligvis trække på den.

eksempler på dataetiske problemstillinger, og diskutere hvordan man systematisk kan identificere dem.

Analysen kombinerer et litteraturstudie med perspektiveringer til forskningslitteraturen om dataetik specifikt og relevante dele af den moderne forskning i normativ og anvendt etik mere generelt.

Det første afsnit definerer og diskuterer hvad en dataetisk problemstilling er. Vi introducerer i den forbindelse forskellen på trivielle dataetiske problemstillinger og dataetiske dilemmaer. I de følgende afsnit giver vi først en række korte eksempler på trivielle dataetiske problemstillinger, og derpå en række mere udførlige eksempler på dataetiske dilemmaer.

De dataetiske dilemmaer inkluderer indholdsregulering på digitale platforme, profilering af personfølsomme persondata, bred indsamling og deling af persondata, og algoritmisk forskelsbehandling. Fælles for disse problemstillinger er, at de involverer vanskelige etiske afvejninger af dataetiske værdier og principper.

I analysens sidste afsnit diskuterer vi hvordan Dataetisk Råd fremadrettet kan identificere dataetiske problemstillinger, som det kan være relevant for Rådet at arbejde med. Der findes ikke en etableret metode til identifikation af sådanne problemstillinger, men vi peger på at en kombination af kritisk evaluering af problemstillinger som italesættes i den offentlige debat, og analyse af nye, omfattende eller indflydelsesrige former for databehandling kan danne udgangspunkt for Rådets forsøg på at identificere problemstillinger.

Hvad er en dataetisk problemstilling?

For at identificere og analysere dataetiske problemstillinger, må man først gøre sig klart, hvad en dataetisk problemstilling er. Vi har tidligere defineret dataetik, som et felt af problemstillinger, hvor databehandling i sig selv udfordrer etiske værdier eller principper. Det vil sige, at en dataetisk problemstilling er en situation, hvor en agent behandler data, og databehandlingen i sig selv negativt påvirker noget etisk værdifuldt eller strider mod (andre) etiske grunde til at handle.

Det er indlysende, at en problemstilling kun er *data*etisk, når den vedrører en form for databehandling. Det er også klart, at en problemstilling kun er data*etisk*, når den vedrører etiske værdier og principper. Endelig er det nødvendigt, at det er databehandlingen i sig selv, og ikke noget andet ved agentens handlinger, som udfordrer etiske værdier og principper. Hvis eksempelvis en person køber hælervarer ved at betale digitalt (eksempelvis med Bitcoin), så bør man skelne mellem den etiske vurdering af det at købe hælervarer, som antageligt er etisk problematisk, og den etiske vurdering af at bruge digitale betalingsmidler. Selvom brugen af digitale betalingsmidler er en form for databehandling, så involverer eksemplet således næppe en *dataetisk* problemstilling, fordi det snarere er købet af hælervarer end databehandlingen i sig selv, som udfordrer etiske værdier og principper.

Det er i denne forbindelse vigtigt at holde sig for øje, at databehandling som *udfordrer* etiske værdier og principper ikke nødvendigvis er etisk forkert, samlet set. At en form for databehandling udfordrer etiske værdier og principper betyder kun, at der i hvert fald i én henseende er noget, som etisk taler imod denne form for databehandling. Dataetiske problemstillinger omfatter således både situationer hvor det er tydeligt, at databehandlingen er etisk forkert, og situationer hvor der er grunde både for og imod en form for databehandling, og det derfor er mindre klart, hvordan man bør handle.[47] Enhver dataetisk problemstilling befinder sig på et spektrum mellem hvad vi kan kalde for trivielle dataetiske problemstillinger og dataetiske dilemmaer.

[47] Det er også en mulighed, at der er dataetiske problemstillinger, hvor det er indlysende, at databehandlingen er etisk, for eksempel fordi der ganske vist er én grund der taler imod databehandlingen, men en række langt mere væsentlige grunde som taler for den. Sådanne problemstillinger er antageligt af begrænset interesse for Rådets arbejde, og vi ser bort fra dem her.

Trivielle dataetiske problemstillinger

I den ene ende af spektret findes dataetiske problemstillinger, som er trivielle i den forstand, at databehandling udfordrer etiske værdier eller principper, og derfor er indlysende forkert.

Når databehandling udfordrer en etisk værdi eller et etisk princip, så findes der som minimum én grund, til ikke at behandle data på denne måde. En dataetisk problemstilling kan derfor være triviel eksempelvis fordi der ikke er andre relevante grunde for eller imod databehandlingen. Hvis der er en etisk grund til ikke at handle på en bestemt måde – i dette tilfælde at udføre en bestemt form for databehandling – og ingen andre etiske grunde for eller imod, så er det klart, at vi ikke bør handle på denne måde. Det kan også være tilfældet, hvis der findes en grund til ikke at udføre den pågældende form for databehandling, som tydeligvis vejer tungere end de grunde der findes, til at udføre eller tillade databehandlingen.[48]

Det kan være kontroversielt hvilke eksempler på databehandling som er og ikke er trivielle, men mange vil nok være enige om vurderingen af de følgende fire eksempler.

Uønsket deling af intime billeder

De seneste år har danske medier rapporteret en lang række eksempler på unge, som har oplevet at intime billeder er blevet delt uden deres tilladelse og i modstrid med hvad de ønskede.[49] En ung person kan eksempelvis dele et nøgenbillede med sin kæreste, og forvente eller aftale, at billedet ikke deles med andre, men efterfølgende opleve, at billedet bliver delt videre med venner, skolekammerater eller endda i offentligt tilgængelige internetfora. En sådan deling er en indlysende krænkelse af privatlivet, og det er ikke oplagt at der er plausible etiske værdier eller principper, som kunne tale for at delingen er tilladelig. Som sådan er det trivielt, at personer ikke bør dele intime billeder af personer, som ikke ønsker billedet delt.

[48] I mange tilfælde vil trivielle dataetiske problemstillinger omhandle databehandling, som vi *ikke* bør udføre. For nemheds skyld fokuserer vi her på disse. Det er imidlertid vigtigt at holde sig for øje, at de samme betragtninger kan gøre sig gældende for databehandling, som vi *bør* udføre. Det er muligt, at der findes situationer, hvor der er en etisk grund til at udføre databehandlingen, og ingen andre relevante grunde, og det derfor er trivielt at vi bør udføre databehandlingen. I sådanne situationer kan det være etisk problematisk *ikke* at udføre databehandlingen.

[49] Et meget prominent eksempel er den såkaldte "Umbrella-sag", hvor flere end 1000 personer blev sigtet for besiddelse og deling af videoer af en 15-årig pige med krænkende seksuelt indhold.

Usikker opbevaring af vigtige persondata

En anden type problemstilling, som har været illustreret de seneste år, er usikker opbevaring af vigtige persondata. Opbevaring af persondata er usikker, når den risikerer utilsigtet spredning af data eller ulovlig adgang til data.[50] Det kan eksempelvis være et offentligt register eller en virksomheds brugerdata, som ikke i tilstrækkelig grad bliver beskyttet, og som hackere derfor får adgang til. En sådan usikker opbevaring af vigtige persondata er åbenlyst i strid med dataetiske krav til sikkerhed, og medfører en alvorlig risiko for krænkelse af personers privatliv. Omvendt er det er ikke klart, at der findes etiske værdier eller principper, som kan tale for, at tillade usikker opbevaring af vigtige persondata. Som sådan er det trivielt, at de relevante agenter bør sørge for sikker opbevaring af vigtige persondata.

Algoritmiske fejlvurderinger

En tredje problemstilling handler om algoritmiske modeller, som bruges til automatiserede beslutningssystemer. En sådan algoritmisk model vurderer en målegenskab, eksempelvis om en arbejdsløs borger lever op til kravene for at kunne modtage en social ydelse, om en bruger på en platform vil være interesseret i et produkt som optræder i en reklame, eller om en kræftknude hos en patient er god- eller ondartet. I nogle tilfælde indgår algoritmens vurdering i et beslutningsgrundlag, som menneskelige agenter bruger til at træffe en beslutning, i andre tilfælde træffes automatisk en beslutning på baggrund af algoritmens vurdering. Algoritmiske modeller kan have høj kvalitet, hvis de er baseret på gode data og udviklet forsvarligt. Omvendt kan en model også have lav kvalitet, hvis de data den er baseret på ikke er gode, eller modellen ikke er udviklet forsvarligt. Lav kvalitet betyder i denne sammenhæng, at modellens vurderinger ofte er fejlagtige. I nogle tilfælde vil en fejlagtig beslutning have uvæsentlige konsekvenser. En algoritme som fejlagtigt vurderer, at en bruger vil være interesseret i et produkt, og derfor viser en relateret reklame, gør i de fleste tilfælde begrænset skade. Men når der er tale om vigtige beslutninger, som har væsentlig indflydelse på personers liv – for eksempel diagnosticeringen af en potentielt livstruende sygdom, eller tildelingen af en social ydelse til en borger med skrøbelig økonomi – så kan fejlagtige

[50] I Danmark har der i de senere år været flere tilfælde, hvor følsomme persondata har været usikkert opbevaret. Eksempelvis lækket af 1,2 millioner CPR-numre, og tyveriet af en computer med data om 20.000 borgere i Gladsaxe kommune.

vurderinger have alvorlige konsekvenser.[51] Udfordringen med algoritmiske fejlvurderinger i tilknytning til vigtige beslutninger har da også været et tilbagevendende kritikpunkt, i særdeleshed i de situationer hvor det har været vanskeligt for berørte personer, at få indsigt i og udfordre beslutningen.[52] Dataetisk betragtet udfordrer et automatiseret beslutningssystem, som begår alvorlige fejl i tilknytning til vigtige beslutninger principper om godgørenhed og skade. Hvis fejlene er så alvorlige, at anvendelse af modellen gør mere skade end gavn, er det vanskeligt at se hvilke etiske værdier og principper, som skulle kunne tale for anvendelsen af den algoritmiske model. I sådanne tilfælde er det trivielt, at agenter ikke bør benytte den algoritmiske model til automatiserede beslutningssystemer.

Dataetik-vask

Den fjerde problemstilling knytter sig til den dataetiske bevidsthed som ofte fremhæves som vigtig i diskussionen af dataetik. Mange virksomheder, organisationer og myndigheder har også udviklet og offentliggjort dataetiske principper og retningslinjer.[53] I nogle tilfælde er sådanne agenter imidlertid blevet kritiseret for, at de dataetiske refleksioner er for abstrakte eller stiller for lave krav til etisk databehandling.[54] I den forbindelse har kritikere rejst bekymringen for, at der kan være tale om "dataetik-vask"[55]: dataetiske refleksioner og principper, som fortrinsvis eller udelukkende har til formål, at få agenten til at fremstå dataetisk ansvarlig overfor andre, for eksempel brugere, myndigheder eller den bredere offentlighed. I værste fald kan dataetik-vask støtte den

[51] De seneste år flere eksempler været bredt debatteret. I Storbritannien måtte regeringen i sommeren 2020 annullere anvendelsen af en algoritme til vurdering af uddannelsessøgende i England, efter massive protester over at algoritmen systematisk undervurderede store grupper studenter. I 2016 deporterede den britiske stat knap 36.000 internationale studerende, som var blevet vurderet til at have snydt ved deres sprogprøver. I 2019 påpegede kritikere, at den algoritme til stemmegenkendelse, som var blevet anvendt til at identificere svindlerne, var fejlbehæftet. Kritikere fremfører, at op til 7.000 studerende kan være blevet uretmæssigt udvist.

[52] Prominente populærvidenskabelige eksempler inkluderer Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. New York, St. Martin's Press; O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, Crown/Archetype; Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, Massachusetts, Harvard University Press.

[53] Prominente eksempler inkluderer New Zealands charter for udvikling og anvendelse af algoritmer, EU-kommissionens charter for etisk anvendelse af AI i retsvæsenet, og ITI's AI policy principper.

[54] Se Wagner, B. (2018). "Ethics as an Escape from Regulation: From Ethics-Washing to Ethics-Shopping?"; Hao, K. (2019). "In 2020, let's stop AI ethics-washing and actually do something."; Bietti, E. (2020). "From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*: 210–219.

[55] Begrebet dataetik-vask trækker på det engelske begreb "greenwashing", som især bruges om handlinger og politikker, der forsøger at fremstå mere miljøvenlige, end de egentlig er, og det afledte begreb "ethics-washing", som især bruges om handlinger og politikker, der forsøger at fremstå mere etiske, end de egentlig er.

misforståelse, at agenten behandler data på en etisk forsvarlig måde, når dette ikke er tilfældet. En anklage om dataetik-vask er mest plausibel i de tilfælde hvor en agents dataetiske refleksioner er tilrettelagt således, at de ikke kan føre til konkrete konklusioner om, hvordan agenten bør behandle data. Det vil ofte være kontroversielt at kategorisere en agents dataetiske refleksioner som dataetik-vask. Det skyldes helt banalt, at agenter har en interesse i at afvise anklagen både når den er grundløs og berettiget. De fleste vil imidlertid kunne blive enige om, at det er eller ville være etisk problematisk at foretage dataetik-vask. Det skyldes at dataetik-vask udfordrer den etiske værdi om (reel) dataetisk bevidsthed, og kan hjælpe agenten med at undgå at handle i overensstemmelse med andre dataetiske værdier og principper. Omvendt er det vanskeligt at pege på en dataetisk værdi eller et dataetisk princip, som kunne tale for dataetik-vask. Samlet er det således trivielt, at agenter ikke bør bedrive dataetik-vask.

Det vil ofte være overflødigt at underkaste trivielle dataetiske problemstillinger en omfattende analyse, fordi en problemstilling netop kun er triviel, når det er indlysende, hvordan agenter bør eller ikke bør behandle data. Udfordringen ved en triviel dataetisk problemstilling vil typisk snarere være praktisk: hvad stiller vi op, for at begrænse eller forhindre de former for databehandling, som er etisk problematiske? Hvordan begrænser eller forhindrer vi, for eksempel, at unge deler intime billeder, som personer på billedet ikke ønsker delt? På denne måde kan en triviel dataetisk problemstilling godt være afsæt for vigtige analyser og værdifuldt arbejde – blot vil sådanne analyser typisk ikke være *dataetiske* analyser.

Dataetiske analyser er derimod afgørende for de dataetiske dilemmaer, som vi nedenfor diskuterer.

Dataetiske dilemmaer

I det foregående afsnit har vi behandlet trivielle dataetiske problemstillinger, som er karakteriseret ved, at de nok udfordrer etiske værdier og principper, men at det er indlysende hvordan agenter bør eller ikke bør behandle data. Det kan eksempelvis skyldes, at der er en klar grund til ikke at udføre en form for databehandling, og ingen oplagte grunde, som kunne tale for at udføre denne type databehandling.

I den modsatte ende af spektret finder vi dataetiske dilemmaer. Et dataetisk dilemma er en dataetisk problemstilling, som er karakteriseret ved, at databehandlingen vedrører flere dataetiske

værdier eller principper, som trækker i forskellige retninger.[56] I sådanne situationer er det ofte uklart hvordan man bør handle. Skal man følge det ene princip, som taler for at behandle data på én måde? Eller det andet princip, som taler for at behandle data på en anden måde? For at tage stilling til et dataetisk dilemma er man nødt til at afklare, hvordan man skal afveje hensynet til de forskellige værdier og principper som er på spil. I et dataetisk dilemma er det derfor nødvendigt at analysere problemstillingen, for at vurdere hvilket synspunkt som støttes af de mest solide begrundelser (se Dataetisk Råds analyse "Dataetisk teori og metode"). Kun på denne måde kan man tage velovervejet stilling til, hvordan agenter bør handle i den pågældende type situation. Dataetiske dilemmaer er derfor særligt relevante for Dataetisk Råds arbejde.

Som vi indledningsvis har nævnt findes der ikke en etableret oversigt over dataetiske problemstillinger. Det skyldes at databehandling dækker over en lang række forskellige typer handlinger, at der ikke er konsensus om et sæt af dataetiske værdier og principper, og at dataetik først de seneste år er blevet genstand for fokuseret opmærksomhed i forskningen såvel som den offentlige debat.

Analysens litteraturstudie har imidlertid identificeret en række centrale eksempler på dataetiske dilemmaer. Disse kan tjene som udgangspunkt for videre arbejde og identifikation af andre væsentlige dataetiske dilemmaer. Nedenfor skitserer vi kort tre dataetiske dilemmaer, inden vi i det sidste afsnit går mere i dybden med et fjerde.

Indholdsregulering på digitale platforme

Digitale platforme, hvor brugere kan dele indhold og skrive til og med hinanden, er blevet en væsentlig del af den danske populærkultur, og vigtige fora for den offentlige debat i Danmark. I takt med at datadeling og diskussion på sådanne platforme er blevet en større del af mange personers hverdag er visse af de udfordringer, som vi også kender fra andre sammenhænge begyndt at manifestere sig, samtidig med at de digitale platformes særlige karakter har rejst nye udfordringer. Tre centrale udfordringer knytter sig til den særlige rolle, som de digitale platforme har, når de

[56] Begrebet "dilemma" bruges ofte i den betydning, at man befinder sig i en situation, hvor man er tvunget til at vælge mellem to alternativer, som begge er utilfredsstillende. Vi bruger her "dilemma" i den lidt bredere forstand, at der både er etiske grunde som taler for og etiske grunde som taler imod en form for databehandling.

udvælger hvilket indhold brugere skal have mulighed for at dele, og hvilket af andre brugeres indhold, den enkelte bruger skal vises.[57]

Den første udfordring er, at brugere har mulighed for at sprede misinformation på digitale platforme. Spredning af sådan misinformation kan i visse tilfælde skade etiske værdier. To meget omtalte eksempler er politisk misinformation som påvirker demokratiske valg, og medicinsk misinformation, som påvirker hvordan personer forholder sig til alvorlige helbredsrisici, for eksempel Coronavirus-pandemien.

Den anden udfordring er, at brugere har mulighed for at ytre sig truende, nedsættende og stigmatiserende, både overfor andre brugere konkret og overfor sårbare befolkningsgrupper generelt. Sådan hadtale kan både have den konsekvens, at nogle brugere kun kan deltage i den offentlige debat ved at betale en psykologisk pris, og styrke stigmatisering og forskelsbehandling af sårbare befolkningsgrupper i andre sammenhænge.

Endelig er en tredje udfordring, at de digitale platformes filtrering af indhold betyder, at mange brugere kan havne i såkaldte filterbobler. En filterboble er karakteriseret ved, at det indhold brugeren præsenteres for fortrinsvis er indhold, som bekræfter brugerens forforståelser og værdier. Derved ændres en fælles offentlig debat til adskilte debatter som foregår indenfor mindre fællesskaber. Samtidig kan den gentagne bekræftelse af eksisterende forståelser og værdier, og fraværet af afvigende eller udfordrende indhold, føre til radikalisering af synspunkter og værdier.

Overfor hensynene til at fremme en fælles demokratisk samtale, inklusion af sårbare befolkningsgrupper og korrekt information af offentligheden står for eksempel den demokratiske værdi af ytringsfrihed og hensynet til den enkeltes autonomi. Ved eksempelvis at begrænse mulighederne for at dele indhold, begrænse visning af indhold, eller markere indhold med advarsler, griber en digital platform afgørende ind, i brugernes mulighed for frit at dele indhold og ytringer. Trods det vil mange være enige om, at der er eksempler på indhold på digitale platforme, som går klart over grænsen, og hvor virksomheder og myndigheder derfor legitimt kan reagere ved for

[57] Se Taddeo, M., L. J. S. Floridi and E. Ethics (2016). "The debate on the moral responsibilities of online service providers." *Science and Engineering Ethics* **22**(6): 1575-1603; Berkey, B. (2017). "Business Ethics and Free Speech on the Internet." *Philosophia* **45**(3): 937-945; Alfano, M., J. A. Carter and M. Cheong (2018). "Technological Seduction and Self-Radicalization." *Journal of the American Philosophical Association* **3**: 298-322.

eksempel at blokere indholdet. Men der vil også være mange tilfælde, hvor der kan være udbredt uenighed om, hvorvidt et konkret stykke indhold er gået over grænsen eller ej. Sådan uenighed kan dels skyldes, at der kan være forhold ved det enkelte tilfælde, som er vanskelige at vurdere. Men det vil også skyldes, at der ikke findes en bredt accepteret og præcis forståelse af, hvor grænsen går mellem indhold, som de digitale platforme bør tillade uhindret deling af, og indhold som de bør begrænse deling af.

Profilering af følsomme persondata

Både virksomheder og myndigheder bruger i stigende grad algoritmiske modeller til at profilere personer. Profilering betyder, at modellen statistisk vurderer en bestemt egenskab ved en person på baggrund af information om personens andre egenskaber.

Et enkelt eksempel på profilering er forsikringsselskabers statistiske analyser af kunder, som skal fastslå kundens risikoprofil. Er en potentiel kunde, som ønsker at købe en bilforsikring, en forsigtig chauffør som har meget lav risiko for at blive involveret i uheld, eller en skødesløs og dumdristig chauffør, som har høj risiko? Risikoprofilen er indlysende relevant, hvis forsikringsselskabet skal fastsætte forsikringspræmien så den modsvarer kundens risiko. Selskabet kan forsøge statistisk at vurdere chaufførens risikoprofil, for eksempel på baggrund af hvor mange uheld chaufføren har været involveret i, chaufførens alder og køn, samt hvor længe og hvor meget chaufføren har kørt bil. Ved at analysere hvordan hver af disse egenskaber i andre tilfælde har påvirket chaufførers risiko for uheld, kan profileringen matematisk vurdere risikoprofilen for en konkret person med en bestemt kombination af egenskaber.

Selvom profilering i nogle sammenhænge er blevet anvendt i årtier, så er brugen af profilering vokset betragteligt de seneste år. Stadigt flere agenter har fået mulighed for at anvende algoritmisk profilering på stadigt flere områder. Det skyldes en kombination af bedre adgang til stadigt større mængder data, og bedre muligheder for at udvikle algoritmiske modeller til profilering med maskinlæring. Brugen af profilering rejser imidlertid flere dataetiske problemstillinger. Én sådan problemstilling er, at profilering i nogle tilfælde kan vurdere egenskaber, som udgør følsomme

persondata (se også diskussionen af forskelsbehandling i automatiserede beslutninger nedenfor). I sådanne tilfælde udfordrer profileringen personers privatliv.[58]

To meget omtalte eksempler på at profilering kan udfordre privatlivet er den amerikanske supermarkedskæde Targets brug af algoritmisk kundesegmentering og en algoritmisk model udviklet af forskere på Stanford University til profilering af brugere af et datingsite. I det første eksempel profilerede Target en ung kvinde, og vurderede at hun med stor sandsynlighed var gravid. Vurderingen var baseret på en række isoleret set uskyldige informationer, såsom hendes køn, alder, købshistorik, og internet-søgemønstre. Algoritmens vurdering foranledigede automatisk at Target sendte kvinden reklamer målrettet gravide med posten. Derved afsløredes den hjemmeboende kvindes graviditet overfor hendes forældre.[59] I det andet eksempel trænede forskere en algoritmisk model, som profilerede personer på et datingsite baseret alene på vurdering af deres ansigter i profilbilleder. Algoritmen viste sig at være i stand til at identificere de pågældende personers seksualitet (homoseksuel/heteroseksuel) med stor pålidelighed.[60]

Profilering kan i mange tilfælde tjene legitime formål. Det kan være relevant for både personen selv, myndigheder, virksomheder, og samfundet mere bredt, at foretage statistiske vurderinger af, om en person besidder en bestemt egenskab. Men i de tilfælde hvor information om den pågældende egenskab udgør følsomme persondata, må de hensyn som taler for profilering afvejes mod hensynet til at beskytte personers privatliv.

Bred indsamling og deling af persondata

En vigtig del af baggrunden for, at vi oplever stadigt nye og mere indflydelsesrige former for databehandling er at det i mange sammenhænge er blevet stadigt nemmere bredt at indsamle store sæt persondata. Ofte vil agenter have en interesse i bred indsamling af persondata, også når disse data ikke umiddelbart skal anvendes til et konkret og specifikt formål. Persondata kan

[58] For et overblik, se Van Den Hoven, J., M. Blaauw, W. Pieters and M. Warnier (2020). "Privacy and Information Technology." I: Zalta, E. (Ed.), *Stanford Encyclopedia of Philosophy*.

[59] Duhigg, C. (2012). "How Companies Learn Your Secrets." *The New York Times Magazine*. Det er værd at bemærke, at kritikere har hævdet, at historien er for god til at være sand. Se Piatetsky, G. (2014). "Did Target Really Predict a Teen's Pregnancy? The Inside Story." *KDnuggets*; Fraser, C. (2020) "Target didn't figure out a teenager was pregnant before her father did, and that one article that said they did was silly and bad." *Medium*.

[60] Wang, Y. and M. Kosinski (2018). "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images." *Journal of personality and social psychology* **114**(2): 246-257.

eksempelvis indsamles fordi agenten forventer, at i hvert fald nogle data på et senere tidspunkt, eventuelt i kombination med data fra andre kilder, kan vise sig at være nyttige.

To almindelige eksempler på brede indsamlinger af persondata er indsamling af brugerdata på internettet, hvor "cookies" kan indsamle data om for eksempel hvilke hjemmesider en bruger besøger, og såkaldt tele-logging, hvor telefonselskaber indsamler metadata, såsom mobiltelefoners placering og hvilke andre telefoner den enkelte mobil har ringet eller skrevet til.

Det centrale spørgsmål i forbindelse med bred indsamling af persondata er om databehandlingen er proportionel. Det vil sige, om den gavn databehandlingen gør, er tilstrækkeligt stor til at legitimere den omkostning dataindsamlingen har.[61]

Kritikere har således fremført, at indsamlingen af data reducerer personers privatliv. Når eksempelvis tele-logging registrerer en mobiltelefons placering, og mange personer har deres mobil på sig på eller i umiddelbar nærhed på alle tidspunkter af døgnet, så bliver det i princippet muligt at foretage en meget detaljeret kortlægning af, hvor den pågældende person har opholdt sig. Hvis privatliv er en etisk værdi, og mange menneskers privatliv på denne måde reduceres væsentligt, så skal der ifølge et proportionalitetsprincip være tilsvarende tungtvejende grunde til at foretage databehandlingen. Kritikere vil i den forbindelse kunne fremføre, at dataindsamling som ikke har et konkret og specifikt formål gør det vanskeligt at pege på sådanne grunde.

Fortalere for indsamling af brede sæt persondata fremfører modsat ofte, at indsamlingen af data kan være værdifuld i visse situationer. Et almindeligt eksempel er at tele-logging kan vise sig at være afgørende i forbindelse med politiets efterforskning af særligt grove forbrydelser. Dertil kan fortalere

[61] Se Milaj, J. (2016). "Privacy, surveillance, and the proportionality principle: The need for a method of assessing privacy implications of technologies used for surveillance." *International Review of Law, Computers & Technology* **30**(3): 115-130; Rønn, K. V. and K. Lippert-Rasmussen (2020). "Out of Proportion? On Surveillance and the Proportionality Requirement." *Ethical Theory Moral Practice* **Online first**; Thomsen, F. K. (2020). "The Teleological Account of Proportional Surveillance." *Res Publica* **26**: 373-401. Det er vigtigt i denne forbindelse at skelne skarpt mellem det *juridiske* spørgsmål om, hvorvidt dataindsamling er proportionel, eksempelvis jf. GDPR-lovgivningens krav til proportionalitet, og det *etiske* spørgsmål om, hvorvidt dataindsamling er proportionel. De to spørgsmål kan naturligvis minde om hinanden, men man kan ikke slutte fra det ene til det andet: dataindsamling kan være uetisk, selvom den er lovlig (hvilket i nogle tilfælde kan udgøre en begrundelse for at stramme eksisterende lovgivning), og ulovlig selvom den er etisk (hvilket i nogle tilfælde kan udgøre en begrundelse for en mindre stram regulering).

hævde, at indsamlingen af data udgør en begrænset reduktion af privatliv i de tilfælde, hvor de pågældende data ikke behandles yderligere.[62]

I mange tilfælde kan store sæt af uspecifikke data også være interessante for at identificere mønstre i data. Eksempelvis kan sådanne data være interessante for økonomer, sociologer og epidemiologer, som ønsker at identificere mønstre i sociale-, forbruger- eller sundhedsdata. En undersøgelse af data forudsætter naturligvis deling af de pågældende data. Når store datasæt deles vil det indlysende være en reduktion af personers privatliv, hvis de kan identificeres i datasættet. En persons privatliv reduceres antageligt jo flere personer, som har adgang til data om personen, og ved deling får nye agenter netop sådan adgang til information om personerne. For at beskytte personers privatliv foretages derfor normalt pseudonymisering af datasættet. Deling af pseudonymiserede datasæt rejser imidlertid en beslægtet problematik: risikoen for re-identifikation.

Pseudonymisering betyder, at de data som umiddelbart kan identificere personer fjernes fra datasættet. Det kan eksempelvis være navn, CPR-nummer og telefonnummer. For at bevare koblingen mellem data og personer tildeles hver person i stedet et pseudonym, for eksempel et tilfældigt genereret nummer.

Pesudonymisering indebærer imidlertid en risiko for, at personer kan reidentificeres. Det vil sige, at deres identitet kan fastslås ved at kombinere informationer som optræder i datasættet med information fra andre datakilder. Studier har vist, at det ofte er muligt at reidentificere pseudonymiserede personer baseret på let tilgængelig information, for eksempel alder, køn og postnummer.[63] I et berømt eksempel reidentificerede den amerikanske forsker Latanya Sweeney staten Massachusets guvernør William Weld i et pseudonymiseret datasæt med sundhedsdata. For

[62] Nogle teorier om privatliv hævder, at en persons privatliv reduceres når andre personer tilgår (eng. "access" (*vb.*)) de relevante data. En mulig konsekvens af adgangsteorien er derfor, at en rent digital indsamling af data i mindre grad eller slet ikke reducerer privatliv, om end den *muliggør* en efterfølgende reduktion af privatliv. Se Macnish, K. (2012). "Unblinking eyes: the ethics of automating surveillance." *Ethics and Information Technology* **14**(2): 151-167.

[63] Sweeney, L. (2000). "Simple Demographics Often Identify People Uniquely." *Data Privacy Working Paper*, Carnegie Mellon University. **3**; Golle, P. (2006). "Revisiting the uniqueness of simple demographics in the US population." *Proceedings of the 5th ACM workshop on Privacy in electronic society***:** 77–80.

at demonstrere risikoen ved offentliggørelse af pseudonymiserede data sendte hun guvernørens personlige helbredsoplysninger til hans kontor.[64]

Der findes i dag en lang række tekniske metoder til at beskytte personers privatliv i pseudonymiserede datasæt. Fælles for disse metoder er imidlertid, at bedre beskyttelse har tendens til at reducere kvaliteten af data. Det betyder at der, ligesom ved indsamlingen af data, er en uundgåelig afvejning imellem hensynet til beskyttelse af personers privatliv og hensynet til de værdifulde formål, som delingen af data kan tjene.

Bias, fairness og forskelsbehandling i automatiserede beslutningssystemer

Det fjerde eksempel på et dataetisk dilemma vedrører forskelsbehandling i automatiserede beslutningssystemer. Et automatiseret beslutningssystem er et system, hvor en algoritmisk model statistisk vurderer værdien for en målegenskab. Det kan, med de eksempler vi tidligere har brugt, være om en person vil være interesseret i reklamen for et produkt, om en kræftknude er godartet, eller om en ansøger er berettiget til en social ydelse. Vurderingen ligger efterfølgende til grund for en beslutning, enten som information til en menneskelig beslutningstager, eller som udgangspunkt for en rent automatisk beslutning.

Vi har ovenfor behandlet den dataetiske problemstilling, som knytter sig til, at algoritmiske modellers vurderinger kan være fejlagtige, og at dette kan have alvorlige konsekvenser for de personer som vurderes. Der findes imidlertid en beslægtet problematik, som handler om at den algoritmiske model i nogle tilfælde systematisk kan forskelsbehandle udsatte grupper. Denne problemstilling diskuteres ofte som spørgsmålet om bias og fairness i algoritmiske modellers vurderinger.[65] I dette afsnit illustrerer vi hvordan man dataetisk kan analysere problemstillingen.

En algoritmisk model som bruges til at træffe beslutninger kan forskelsbehandle grupper på to forskellige måder. For det første kan en model *direkte* forskelsbehandle en gruppe. Det gør den, hvis

[64] Ohm, P. (2010). "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." *UCLA Law Review* **57**: 1701-1777; Sweeney, L. (2000). "Simple Demographics Often Identify People Uniquely." *Data Privacy Working Paper*, Carnegie Mellon University.

[65] Se Barocas, S. and A. D. Selbst (2016). "Big Data's Disparate Impact." *California Law Review* **104**(3): 671-732; Kleinberg, J., J. Ludwig, S. Mullainathan and C. R. Sunstein (2019) "Discrimination in the Age of Algorithms." *arXiv e-prints;* Chouldechova, A. and A. Roth (2018) "The Frontiers of Fairness in Machine Learning." *arXiv e-prints;* Binns, R. (2018). "Fairness in Machine Learning: Lessons from Political Philosophy." *Journal of Machine Learning Research* **81**: 1-11.

modellen bruger medlemskab af gruppen som variabel, og medlemskab af gruppen derved påvirker modellens vurderinger. Hvis eksempelvis en model forsøger at vurdere om en person er berettiget til en social ydelse, og bruger data om personers køn til at justere sandsynligheden op for mænd og ned for kvinder, så forskelsbehandler modellen de to grupper direkte.

Direkte forskelsbehandling udgør i nogle tilfælde ulovlig direkte diskrimination. Det kan være tilfældet hvis en model direkte forskelsbehandler særligt beskyttede grupper, for eksempel ved at forskelsbehandle på baggrund af køn, etnicitet, religion, eller handicap. Af samme grund undgår udviklere ofte at benytte medlemskab af beskyttede grupper som variable i modeller.

Debatten om bias og fairness fokuserer derfor typisk på den problemstilling, at en model kan *indirekte* forskelsbehandle beskyttede grupper. Sådan indirekte forskelsbehandling er vigtig, fordi den kan forekomme selvom udvikleren har sikret at modellen ikke direkte forskelsbehandler, og fordi den i mange tilfælde kan være vanskelig både at opdage og forhindre.

En model forskelsbehandler grupper indirekte, hvis den har tendens til at vurdere grupperne forskelligt *selvom* den ikke direkte forskelsbehandler dem. Det kan eksempelvis forekomme, fordi modellen benytter andre variable, som hænger statistisk sammen med medlemskab af pågældende grupper. Med et meget enkelt eksempel, så kan en model have tendens til at behandle mænd og kvinder forskelligt hvis den bruger personers højde som variabel, fordi mænd i gennemsnit er højere end kvinder. I praksis benytter en model typisk mange forskellige egenskaber som variable – ind imellem flere hundrede eller tusinde forskellige egenskaber. I sådanne tilfælde kan hver enkelt egenskab have meget lille effekt, og alligevel til sammen give modellen meget stærk tendens til at forskelsbehandle grupper.

Kritikere har især fremført to anklager mod sådan indirekte forskelsbehandling. For det første, at de sammenhænge som modellen er baseret på i nogle tilfælde er et resultat af tidligere forskelsbehandling, og indirekte forskelsbehandling som afspejler disse sammenhænge er etisk problematisk. Eksempelvis kan et uddannelsessystem bruge en algoritmisk model til at vurdere hvordan ansøgere til videregående uddannelse vil klare sig på uddannelsen. En sådan vurdering kan for eksempel støtte eller træffe beslutningen om hvilke ansøgere som skal optages. Et oplagt udgangspunkt for sådanne vurderinger vil være ansøgerens tidligere eksamensresultater, men hvis

medlemmer af en bestemt gruppe i gennemsnit har lavere karakterer end andre *fordi* de er blevet udsat for diskrimination, så vil modellen gentage eller endda forstærke denne diskrimination, når den bruger eksamensresultaterne som grundlag for vurderingen.

I nogle tilfælde kan en løsning være at rense data for den indflydelse, som tidligere diskrimination har haft. Det vil oplagt være tilfældet, hvis en gruppes lavere eksamensresultater alene skyldes at de er blevet diskrimineret i eksamenssituationen. Det er imidlertid et åbent spørgsmål hvordan man skal definere og måle indflydelsen af tidligere diskrimination, og derfor i praksis ofte vanskeligt at rense data.

Den anden anklage hævder at det kan være etisk problematisk indirekte at forskelsbehandle grupper uanset om de relevante forskelle mellem grupperne skyldes tidligere diskrimination. Det kan dels være tilfældet, når forskellene ikke skyldes diskrimination, som i eksemplet med forskelle i mænds og kvinders højde. Det kan også være tilfældet i situationer hvor forskellene er relevante *selvom* de skyldes diskrimination. Forskellen i to gruppers eksamensresultater kan for eksempel afspejle, at den ene gruppe reelt har gennemsnitligt ringere kvalifikationer *fordi* gruppen er blevet diskrimineret i hele det tidligere uddannelsesforløb. I en sådan situation virke urimeligt at stille krav om, at modellens data renses så den vurderer hvilke kvalifikationer ansøgere *ville have haft*, hvis de ikke var blevet udsat for diskrimination, fordi dette kan føre til optagelse af personer, som reelt ikke har de fornødne kvalifikationer til at gennemføre uddannelsen. Spørgsmålet er hvad det i en sådan situation kræver, at behandle de pågældende grupper ligeligt/fair?

Den foreløbige debat om bias og fairness i (indirekte) algoritmisk forskelsbehandling har fokuseret på fordele og ulemper ved tre forskellige måder at forstå dét, at behandle grupper lige:

- Demografisk paritet: Fordelingen af vurderinger afspejler gruppernes demografiske andele. Hvis eksempelvis en model vurderer at 10% af personerne i én gruppe har en målegenskab, så kræver demografisk paritet at modellen også vurderer at 10% af personerne i andre grupper har målegenskaben.
- Præcisionsparitet: Modellens vurderinger er lige præcise for alle grupper. Hvis eksempelvis en model laver korrekte vurderinger i 90% af tilfældene, når den vurderer personer fra én

gruppe, så kræver præcisionsparitet at modellen også laver korrekte vurderinger i 90% af tilfældene, når den vurderer personer fra andre grupper.

- Fejltypeparitet: Modellen laver de samme typer fejl, når den laver fejl, for alle grupper. En algoritmisk model vurderer en målegenskab, for eksempel om en kræftknude er godartet eller ondartet. Fejltypeparitet kan være vigtig, fordi det kan være afgørende, hvilken slags fejl modellen laver. En model, som vurderer kræftknuder, laver fejl både når den vurderer at en godartet knude er ondartet, og når den vurderer at en ondartet knude er godartet. Vi vil helst undgå begge typer fejl, men hvis vi skal vælge, så er det måske alligevel vigtigere, at undgå de fejl, hvor ondartede knuder bliver fejlvurderet som godartede. Tilsvarende kan en model forskelsbehandle to grupper, selvom den har lige høj præcision for de to grupper, hvis den har tendens til at lave den ene slags fejl for den ene gruppe, og den anden slags fejl for den anden gruppe.

Den dataetiske analyse af fairness og bias rejser to vigtige spørgsmål om disse tre typer forskelsbehandling. For det første, så har det vist sig, at det ofte vil være matematisk umuligt på samme tid at behandle forskellige grupper lige i alle de tre betydninger af ligebehandling, som vi ovenfor har skitseret.[66] I sådanne situationer er vi tvunget til at beslutte, hvilken type ligebehandling vi etisk bør prioritere.

For det andet, så bør vi i den forbindelse afklare *hvorfor* det er vigtigt at undgå visse typer forskelsbehandling. På et generelt niveau kan man skelne mellem tre typer forklaringer på, hvorfor diskrimination er etisk problematisk:

- Diskrimination er etisk problematisk fordi det skader de diskriminerede.
- Diskrimination er etisk problematisk fordi det forøger uretfærdige uligheder mellem de diskriminerede og andre.
- Diskrimination er etisk problematisk fordi det ikke respekterer de diskriminerede.[67]

[66] Se Kleinberg, J., S. Mullainathan and M. Raghavan (2016) "Inherent Trade-Offs in the Fair Determination of Risk Scores." *arXiv e-prints;* Lipton, Z. C., A. Chouldechova and J. McAuley (2018). Does mitigating ML's impact disparity require treatment disparity? *32nd Conference on Neural Information Processing Systems;* Chouldechova, A. (2017). "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments." *Big Data* **5**(2): 153-163.
[67] For overblik, se Altman, A. (2015). "Discrimination." I: Zalta, E. (Ed.), *Stanford Encyclopedia of Philosophy*; Thomsen, F. K. (2017). "Discrimination." I: Thompson, W.R. (Ed.), *Oxford Research Encyclopedia of Politics*.

Selvom der det seneste årti har været en intens diskussion af sådanne forklaringer, så er der ikke konsensus om hvordan forklaringerne kan overføres til indirekte diskrimination mere generelt eller algoritmisk forskelsbehandling specifikt. [68] En endelig dataetisk analyse af algoritmisk forskelsbehandling forudsætter derfor en yderligere afklaring af disse vigtige etiske spørgsmål.

[68] Se Eidelson, B. (2015). *Discrimination and Disrespect*. Oxford, Oxford University Press; Thomsen, F. K. (2015). "Stealing Bread and Sleeping Beneath Bridges - Indirect Discrimination as Disadvantageous Equal Treatment." *Moral Philosophy and Politics* **2**(2): 299-327; Lippert-Rasmussen, K. (2014). "Indirect Discrimination is Not Necessarily Unjust." *Journal of Practical Ethics* **2**(2): 33-57.

Hvordan identificerer man en dataetisk problemstilling?

I de foregående afsnit har vi defineret dataetiske problemstillinger, og præsenteret en række prominente eksempler på aktuelle dataetiske problemstillinger. Dataetisk Råd arbejder med at analysere og diskutere dataetiske problemstillinger. Et vigtigt spørgsmål for Dataetisk Råds arbejde er derfor hvordan man konkret identificerer en dataetisk problemstilling? I dette sidste afsnit diskuterer vi dette spørgsmål.

Vi har defineret en dataetisk problemstilling som en situation, hvor en agent behandler data, og databehandlingen i sig selv udfordrer etiske værdier eller principper. Det et vilkår for dataetikken, at databehandling spænder over en lang række forskellige handlinger, samt at den hastige udvikling af ny teknologi konstant åbner nye muligheder for databehandling. Det kan derfor være en uoverskuelig opgave, at undersøge hver enkelt form for databehandling, for at vurdere om databehandlingen rejser en dataetisk problemstilling. Det er nødvendigt at fokusere indsatsen for at identificere dataetiske problemstillinger.

Der findes ikke en etableret metode til at identificere etiske problemstillinger, men på et generelt niveau kan man skelne mellem to ofte anvendte måder at fokusere på mulige dataetiske problemstillinger:

- Man kan undersøge databehandling, som bliver italesat som dataetisk problematisk, eksempelvis i den offentlige debat.
- Man kan undersøge særligt relevante former for databehandling, især databehandling som er ny, omfattende eller indflydelsesrig.

Italesatte dataetiske problemstillinger

Den mest enkle metode er at tage fat i databehandling, som er blevet *italesat* som en dataetisk problemstilling. Databehandling italesættes som dataetisk problemstilling, hvis taleren implicit eller eksplicit hævder, at man i den pågældende situation bør databehandle en bestemt måde. Italesættelse af en dataetisk problemstilling kræver således ikke, at taleren eksplicit kalder databehandlingen for etisk problematisk, eller henviser til etiske værdier og principper. Påstanden om, at man bør databehandle på en bestemt måde forudsætter nemlig at der findes begrundelser

for synspunktet, og gode begrundelser for hvordan man bør handle trækker nødvendigvis på etiske værdier og principper.[69]

Groft sagt er al offentlig kritik af databehandling derfor italesættelse af en dataetisk problemstilling. Taleren kan for eksempel kritisere databehandling, som man (ifølge taleren) slet ikke burde foretage, eller omvendt kritisere at man ikke foretager databehandling, som man burde foretage, og endelig hævde at man burde foretage databehandling, men på en anden måde.

Den offentlige debat i Danmark rummer jævnligt eksempler på databehandling, som på denne måde bliver gjort til genstand for kritik. Det er imidlertid nødvendigt kritisk at vurdere de problemstillinger som italesættes. Det skyldes at italesættelsen af en dataetiske problemstilling nok kan tages som udtryk for, at taleren opfatter databehandlingen som dataetisk problematisk. Men deraf følger jo ikke, at databehandlingen faktisk *er* dataetisk problematisk. Eksempelvis kan det vise sig, at opfattelsen af databehandling som dateetisk problematisk kun kan forsvares med henvisning til etiske værdier og principper, som der ikke er gode grunde til at acceptere. I så fald må man afvise, at der er tale om en reel dataetisk problemstilling.

Særligt relevante former for databehandling

Den første metode, som tager udgangspunkt i allerede italesatte dataetiske problemstillinger, er både enkel og velegnet til at fokusere Rådets arbejde på aktuelle samfundsdebatter. Metoden har til gengæld den begrænsning, at den alene kan identificere dataetiske problemstillinger, som allerede er blevet italesat som dataetiske problemstillinger. Metoden medfører således en risiko for, at der findes vigtige dataetiske problemstillinger, som ikke bliver identificeret, simpelthen fordi de ikke allerede er blevet italesat som dataetisk problemstilling. Den første metode kan derfor med fordel suppleres med forsøg på at identificere oversete dataetiske problemstillinger.

Disse forsøg må nødvendigvis fokusere på særligt relevante former for databehandling. Databehandling kan i denne sammenhæng forstås som særligt relevant hvis den er ny, omfattende eller indflydelsesrig.

[69] Det er muligt at fremføre *ugyldige* begrundelser for hvordan man bør handle, som ikke henviser til etiske værdier og principper. Sådanne begrundelser begår den såkaldte naturalistiske fejlslutning.

Hvis databehandlingen er ny er der en større sandsynlighed for, at den er dataetisk problematisk, men at dette ikke er blevet erkendt endnu. Når databehandling har været foretaget i en længere periode, kan man i højere grad forvente, at den ville være blevet kritiseret, hvis den rejste dataetiske problemstillinger.

Når databehandlingen er omfattende eller indflydelsesrig, kan man forvente at eventuelle dataetiske problemstillinger er vigtige. Når databehandlingen har beskedent omfang eller begrænsede konsekvenser kan man omvendt forvente, at problemets omfang vil være beskedent, selv hvis databehandlingen er etisk problematisk.

Særligt relevante former for databehandling rejser ikke nødvendigvis en dataetisk problemstilling. For at rejse en dataetisk problemstilling skal databehandlingen udfordre etiske værdier og principper. Det vil derfor videre være nødvendigt at undersøge hver enkelt særligt relevant form for databehandling, for at vurdere om dette er tilfældet.

BILAG

Litteraturstudier

Bilag 1: Litteraturstudie (definitioner af dataetik)

Litteraturstudiet er foretaget ved søgninger på scholar.google.com, google.com og forskningsdatabasen.dk i september 2020.

Litteraturstudiet benyttede følgende søgeord på hver af de tre databaser:

- "data ethics"
- data ethics
- data ethics definition
- "data ethics is"
- "data Ethics can be defined as"
- big data ethics
- ethics of data
- "dataetik"

Søgningerne genererede forventeligt mange hundrede resultater. Disse resultater er filtreret to gange. Ved de umiddelbare søgninger frasorteredes indlysende irrelevante resultater. De tilbageværende resultater blev indhentet og studeret. Ved læsning af disse resultater blev yderligere en håndfuld resultater sorteret fra som irrelevante.

Litteraturstudiet har identificeret 43 relevante kilder (foruden Dataetisk Råds kommissorium), hvoraf 19 er akademiske kilder, mens 24 er policy-dokumenter og øvrige kilder

Kilderne optræder nedenfor citeret med forfatter, årstal, titel og hyperlink. De centrale passager fra hver kilde er angivet i citat. Kilderne er i den forbindelse inddelt i fem kategorier:

- Eksplicitte definitioner, forskningslitteratur
- Eksplicitte definitioner, policy-dokumenter og øvrige kilder
- Implicitte definitioner, forskningslitteratur
- Implicitte definitioner, policy-dokumenter og øvrige kilder
- Kontekstspecifikke, implicitte definitioner, forskningslitteratur

Kilde	Citat
Eksplicitte definitioner, forskningslitteratur	
L. Floridi & M. Taddeo 2016: *What is data ethics?*	"Data Ethics can be defined as the branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing, and use), algorithms (including AI, artificial agents, machine learning, and robots), and corresponding practices (including responsible innovation, programming, hacking, and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values) [...] The ethics of data focuses on ethical problems posed by the collection and analysis of large dataset and on issues ranging from the use of Big Data in biomedical research and social sciences, to profiling, advertising, and data philanthropy as well as open data." (p. 5)
G. Hasselbach 2019: *Making sense of data ethics. The powers behind the data ethics debate in European policymaking*	"[...] As a point of departure, I define a data ethics of power as an action-oriented analytical framework concerned with making visible the power relations embedded in the "Big Data Society" and the conditions of their negotiation and distribution, in order to point to design, business, policy, social and cultural processes that support a human-centric distribution of power" (p. 2-3) "[...] data ethics can be viewed as a proactive agenda concerned with shifting societal power relations and with the aim to balance the powers embedded in the Big Data Society [...] Here, I use the notion of "Big Data Society" to reflectively position data ethics in the context of a recent data (re)evolution of the "Information Society", enabled by computer technologies and dictated by a transformation of all things (and people) into data formats ("datafication") in order to "quantify the world" to organise society and predict risks" (p. 3-4) "What is data ethics? Currently, the reply is shrill, formally framed in countless statements, documents and mission statements from a multitude of sources, including governments, intergovernmental organisations, consultancy firms, companies, non-governmental organisations, independent experts and academics. But it also emerges when least expected, in "non-allocated" moments of discussion. Information technologies that permeate every aspect of our lives today, from micro work settings to macro economics and politics, are increasingly discussed as "ethical problems" that must be solved. Their pervasiveness sparks moments of ethical thinking, negotiated in terms of moral principles, values and ideal conditions. In allocated or unallocated spaces of negotiation, moments of pause and sense-making, we discuss the values and politics of the business practices, cultures and legal

	jurisdictions that shape them. These spaces of negotiation encompass very concrete discussions regarding specific information technology tools, but increasingly they also evolve into reflections concerning general challenges to established legal frameworks, individuals' agency and human rights, as well as questions regarding the general evolution of society. As one Danish minister said at the launch of a national data ethics expert group: "This is about what society we want"." (p. 13)
N. Mishra 2020: *International Trade Law Meets Data Ethics: A Brave New World*	"Broadly, the article refers to all moral or ethical concerns regarding data as 'data ethics'. This formulation of data ethics aligns with the definition put forth by Floridi and Taddeo, who define data ethics as a 'moral compass' to determine 'good' digital regulation and governance"
Eksplicitte definitioner, policy-dokumenter og øvrige kilder	
Danish Expert Group on Data ethics 2018: *Data for the Benefit of the People : Recommendations from the Danish Expert Group on Data Ethics*	"Data ethics is defined as an active decision and action to ensure that knowledge obtained through data is not used against the legitimate interests of an individual or group. With data ethics, organisations work actively to address data ethics issues in design, innovation and business processes. When we talk about data ethics in relation to companies, responsibility lies beyond what is stated in data protection legislation (e.g. GDPR). Data ethics is about doing the right thing, even when no one is watching" (p. 7)
ODI 2018: *What is data ethics?*	"What is data ethics? "Data ethics is a branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use""
Department for Digital, Culture, Media & Sport 2018: *Guidance: Data Ethics Framework*	"What is data ethics? Data ethics is an emerging branch of applied ethics which describes the value judgements and approaches we make when generating, analysing and disseminating data. This includes a sound knowledge of data protection law and other relevant legislation, and the appropriate use of new technologies. It requires a holistic approach incorporating good practice in computing techniques, ethics and information assurance" (p. 3)
Tranberg et. al. 2018: *DATAETHICS – Principles and Guidelines for*	"What is data ethics? Data ethics is about responsible and sustainable use of data. It is about doing the right thing for people and society. Data processes should be designed as sustainable solutions benefitting first and foremost humans. Data ethics refer and adhere to the principles and values on which human rights and personal data protection laws are

Companies, Authorities & Organisations	based. It's about honest and genuine transparency in data management. To actively develop privacy-by-design and privacy-enhancing products and infrastructures. To treat someone else's personal information as you wish your own, or your children's, treated. Data ethics is the step further than mere compliance with personal data protection laws [...]" (p. 7-8)
Erhvervsstyrelsen: *Hvad er dataetik?*	"Hvad er dataetik? Dataetik handler om ansvarlig og bæredygtig brug af data samt at skabe gennemsigtighed i virksomhedens datahåndtering. Dataetik er altså ikke bare et spørgsmål om at overholde lovgivningen, men handler om at behandle andres data med respekt og gøre det rigtige, selv når ingen kigger"
Forsikring & Pension: *Cool eller creepy? Databrug og dataetiske principper i forsikrings- og pensionsbranchen*	"Dataetik i forsikrings- og pensionsbranchen handler om samspillet mellem etik, innovation og forretning, men for at dataetik skal være mere end en tynd fernis af gode intentioner, er der behov for at sætte fokus på, hvad dataetik betyder i praksis [...] Etik handler om mennesker, og udgangspunktet er, at der er en naturlig vekselvirkning mellem de digitale muligheder og den menneskelige vilje til at tilpasse sig. Den balance prøver principperne [de principper, rapporten præsenterer] at indfange efter bedste evne, men det er også givet, at dataetiske principper aldrig bliver mejslet i granit, og at der kontinuerligt vil være behov for videreudvikling" (p. 5)
Forsikring & Pension 2019: *Databrug og dataetik: Dilemmaer og mulige positioner for forsikrings- og pensionsbranchen*	"For forsikrings- og pensionsbranchen handler dataetik om at finde en balance i de mange de svære dilemmaer. I stedet for at grave skyttegrave, ønsker vi at gå foran i en åben diskussion om, hvad vi skal gøre med den store mængde data, vi alle producerer i dag. Det er uundgåeligt, at svære etiske valg og dilemmaer melder sig. Men de forsvinder ikke ved at ignorere dem. Den bedste måde at håndtere dem på er ud fra en klart defineret dataetisk position"
DANSK IT's arbejdsgruppe for dataetik 2018: *Dataetik: 18 dataetiske anbefalinger fra DANSK IT*	"Hvad forstår vi ved begrebet dataetik? [...] mens vi i DANSK IT ofte er med til at fremhæve de mange positive muligheder, der skabes med digitaliseringen, mener vi også, at det er helt afgørende at italesætte de dilemmaer og udfordringer, der også følger med udviklingen [...] Som alle andre etiske spørgsmål vedrører spørgsmålet om dataetik grundlæggende, hvad der er det gode og det rigtige for det enkelte menneske og for fællesskabet. Det vil altid være en afvejning af forskellige hensyn, og det vil ofte være dilemmafyldt. Men det betyder ikke, at alt kan være lige rigtigt eller lige forkert. Vi må som samfund og som individer tage stilling og etablere nogle normer, der af de fleste mennesker i de fleste situationer opleves som rigtige og retfærdige. Sådanne normer ændrer sig

	over tid, og de bør derfor med mellemrum tages op til overvejelse og debat. Der findes ingen evigtgyldige svar i debatten om dataetik, men der bør findes svar, som har en høj grad af permanens og som kun ændres langsomt" (p. 11-13)
D. Leslie/The Alan Turing Institute 2019: *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*	"AI ethics is a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies" (p. 3)
Implicitte definitioner, forskningslitteratur	
N. Richards & J. King 2014: *Big Data Ethics*	"[...] we call for the development of "Big Data Ethics," a set of four high-level principles that we should recognize as governing data flows in our information society, and which should inform the establishment of legal and ethical big data norms. To advance ethics of big data, four such principles should be paramount [...] What we need are new rules to regulate the societal costs of our new tools without sacrificing their undeniable benefits [...] we argue that a set of four normative values (privacy, confidentiality, transparency, and identity) suggests the beginnings of "Big Data Ethics" to govern data flows in our information society and inform the establishment of legal and ethical big data norms" (p. 395-408)
K. Crawford, K. Miltner & M. Gray 2014: *Critiquing Big Data: Politics, Ethics, Epistemology*	"How can data be gathered without people's knowledge or consent and still meet the ethical obligation to treat people with "justice, beneficence, and respect," as the Belmont Report on human subjects research first outlined in 1978? Scientific research that involves drawing on what is euphemistically known as "passively collected" big data must face difficult questions and develop new ethical frameworks" (p. 4)
A. Zwitter 2014: *Big Data Ethics*	"The speed of development in Big Data and associated phenomena, such as social media, has surpassed the capacity of the average consumer to understand his or her actions and their knock-on effects. We are moving

	towards changes in how ethics has to be perceived: away from individual decisions with specific and knowable outcomes, towards actions by many unaware that they may have taken actions with unintended consequences for anyone. Responses will require a rethinking of ethical choices, the lack thereof and how this will guide scientists, governments, and corporate agencies in handling Big Data" (p. 1)
S. Vallor 2018: *An Introduction to Data Ethics*	"What does a person given access to all that data, or tasked with analyzing it, need to understand about its ethical significance and power to affect a person's life?" (p. 3) "Ethical issues are everywhere in the world of data, because data's collection, analysis, transmission and use can and often does profoundly impact the ability of individuals and groups to live well" (p. 4) "The combination of increasingly powerful but also potentially misleading or misused data analytics, a data-saturated and poorly regulated commercial environment, and the absence of widespread, well-designed standards for data practice in industry, university, non-profit, and government sectors has created a 'perfect storm' of ethical risks. Managing those risks wisely requires understanding the vast potential for data to generate ethical benefits as well" (p. 6)
Filimowicz, M. & Tzankova, V. 2021: *Data Ethics : A survey of key debates and essential principles*	"[...] this project attempts to provide a useful discussion of some central concerns over data use and management, and to frame these concerns with key principles from ethics and technology theory [...] This project also aims to contribute to current data ethics discourses concerned with the level of social change caused by big data, the ethical repercussions that have manifested as a result, and those that lie ahead […] our project aspires to sketch possible answers to the following questions: What are the inherent moral obligations guiding corporate use and management of data, and how can they be improved? What would an ethically acceptable data-ownership model look like, and who should own the data? And, how would a sensible policy framework safeguard individual autonomy and privacy, while guaranteeing fairness to the private sector's for-profit operational model?" (p. 2)
Herschel & Miori 2017: *Ethics & Big Data*	"A discussion of ethics and Big Data is dependent upon how one defines ethics. In general, ethics involves the analysis of conduct that can cause benefit or harm to other people [...] Sound ethical theories share a common property. They enable the individual to make persuasive, logical, and reasoned arguments based on the principles stated by the ethical theory [...] By examining ethical theories, we can better recognize differing perspectives on Big Data-related moral situations, better understand the context and the logic of the arguments being presented, and in so doing

	better evaluate how the intended course of action is or should be justified" (p. 33-35)
Vayena E, Tasioulas J. 2016: *The dynamics of big data and human rights: the case of scientific research*	"If the way we live our lives is undergoing such transformations, are our existing ethical values still relevant? [...] Do we have to choose between big data and the ethical outlook that values such as privacy represent? Or should we turn to new values for guidance? Ultimately, the central question is this: if big data is here to stay, in some sense, what kind of big data society do we want to have and how can we best achieve it?" (p. 2)
White & Ariyachandra 2016: *BIG DATA AND ETHICS: EXAMINING THE GREY AREAS OF BIG DATA ANALYTICS*	"Big data and ethics is a newer area that still needs specified guidelines. Within the next couple of years, it is recommended that specific ethical standards are developed to manage and care for big data [...] As data collection continues at an exponential pace, greater care must be given to the ethics involved in the manipulation and the use of big data for organizational decision making" (p. 5)
Implicitte definitioner, policy-dokumenter og øvrige kilder	
Regeringen 2019: *Dansk National Strategi for Kunstig Intelligens*	"Med National strategi for kunstig intelligens sætter regeringen fire sigtelinjer for, hvordan Danmark skal gå forrest med ansvarlig udvikling og anvendelse af kunstig intelligens. 1. Danmark skal have et fælles etisk grundlag for kunstig intelligens med mennesket i centrum Kunstig intelligens skal udvikles og anvendes inden for gældende lovgivning og med respekt for borgernes rettigheder. Det indebærer, at virksomheder og offentlige myndigheder skal have et stærkt fokus på dataetik, som blandt andet omfatter ansvarlighed, sikkerhed og gennemsigtighed i brugen af kunstig intelligens. Regeringen har som mål, at: ● etiske principper tænkes ind i udviklingen og anvendelsen af kunstig intelligens, så der sikres respekt for individet og dets rettigheder samt for demokratiet ● Danmark skal tiltrække viden og projekter ved at være blandt de bedste i EU på muligheder for at udvikle og anvende ansvarlig kunstig intelligens ● Danmark arbejder aktivt internationalt for, at ansvarlighed bliver et styrende princip for anvendelse og udvikling af kunstig intelligens" (p. 8)

EDPS Ethics Advisory Group 2018: *Towards a Digital Ethics*	"If we accept the idea of a new digital reality, we also accept that it brings with it changing conditions of being human. It invites a new ethical evaluation, a new interpretation of some of the fundamental notions in ethics, such as dignity, freedom, autonomy, solidarity, equality, justice, and trust; and invites us to test the conditions of their validity for the new realities that present themselves in this new age [...] he purpose of digital ethics is not only to account for the present, but also to perform a foresight function" (p. 15) "Digital ethics can be understood from a number of perspectives. A basic distinction is commonly drawn in ethical reflection between normative (or prescriptive) ethics and metaethics [...] In its deliberations the EAG took the consensus-based decision to focus primarily on metaethical questions of digital ethics. Its work thus consisted of considering more general and fundamental questions about what it means to make claims about ethics and human conduct in the digital age, when the baseline conditions of 'human-ness' are under the pressure of interconnectivity, algorithmic decision-making, machine-learning, digital surveillance and the enormous collection of personal data, about what can and should be retained and what can and should be adapted, from traditional normative ethics" (p. 10)
DEK 2019: *Opinion of the Data Ethics Commission*	"Humans are morally responsible for their actions, and there is no escaping this moral dimension. Humans are responsible for the goals they pursue, the means by which they pursue them, and their reasons for doing so. Both this dimension and the societal conditionality of human action must always be taken into account when designing our technologically shaped future. At the same time, the notion that technology should serve humans rather than humans being subservient to technology can be taken as incontrovertible fact" (p. 14) "Given the increase in the volume of personal data being collected and the use of automated methods to process these data for different purposes, one of the main priorities of the Data Ethics Commission is to reconcile the need to protect the individual's fundamental rights and freedoms – including self-determination and integrity – with the need to promote progress, prosperity, the safeguarding of democracy and the shaping of a society that is fit for the future." (p.37)
Independent High-Level Expert Group on Artificial Intelligence set up by the European	"[...] AI systems need to be human-centric, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom. While offering great opportunities, AI systems also give rise to certain risks that must be handled appropriately and proportionately. We now have an important window of opportunity to shape their development. We want to ensure that we can trust the sociotechnical environments in which they are

Commission 2019: *Ethics Guidelines for Trustworthy AI*	embedded. We also want producers of AI systems to get a competitive advantage by embedding Trustworthy AI in their products and services. This entails seeking to maximise the benefits of AI systems while at the same time preventing and minimising their risks [...] even after compliance with legally enforceable fundamental rights has been achieved, ethical reflection can help us understand how the development, deployment and use of AI systems may implicate fundamental rights and their underlying values, and can help provide more fine-grained guidance when seeking to identify what we should do rather than what we (currently) can do with technology [...] These ethical principles can inspire new and specific regulatory instruments, can help interpreting fundamental rights as our socio-technical environment evolves over time, and can guide the rationale for AI systems' development, deployment and use – adapting dynamically as society itself evolves"" (p. 4-11)
The Agency for Digital Italy 2018: *White Paper on Artificial Intelligence at the service of citizens*	"The ethical challenge of the introduction of Artificial Intelligence solutions is represented by the need to respond in a balanced manner to the polarisation of these two visions, integrating innovation and taking into account the effects that this has already had and will continue to have in the development of society, respecting and safeguarding the universally recognised core values. The use of AI based on algorithms of data analysis in decision-making processes related to social, health and judicial issues (such as risk assessment) therefore requires a thorough reflection in terms of ethics and, more broadly, of governance [...] To address these challenges, it may be helpful to follow some general principles" (p. 34-37)
ACM US Public Policy Council 2017: *Statement on Algorithmic Transparency and Accountability*	"Computer algorithms are widely employed throughout our economy and society to make decisions that have far-reaching impacts, including their applications for education, access to credit, healthcare, and employment. The ubiquity of algorithms in our everyday lives is an important reason to focus on addressing challenges associated with the design and technical aspects of algorithms and preventing bias from the onset [...] The use of algorithms for automated decision-making about individuals can result in harmful discrimination. Policymakers should hold institutions using analytics to the same standards as institutions where humans have traditionally made decisions and developers should plan and architect analytical systems to adhere to those standards when algorithms are used to make automated decisions or as input to decisions made by people"
Amsterdam Economic Board (TADA): *TADA – Data Disclosed*	"Data: a promise for life in the city. Data enables us to tackle major problems of modern cities, making them cleaner, safer, healthier… but only as long as people stay in control of the data, and not the other way round. We – companies, government, communities and citizens – see this

manifesto	as a team effort and want to be a leading example for all other digital cities across the globe. To get started, we have come together to set out the following shared principles [...] In rapidly digitizing cities, ethical and responsible use of data is a major challenge – and Amsterdam is no exception. Professionals from the Amsterdam region therefore wrote a manifesto entitled 'Tada – data disclosed'. Government authorities, companies and other organizations from different regions are invited to use and sign the document, showcasing their ambitions to shape a responsible digital city"
Australian Government: *AI Ethics Framework*	"For Australia to realise the benefits of Artificial Intelligence (AI) the public needs to be able to trust it is safe, secure and reliable. To help build trust in AI, we've committed to developing an AI Ethics Framework to guide businesses and governments looking to design, develop and implement AI in Australia. This is part of the Australian Government's commitment to build Australia's AI capabilities [...] You can use our 8 principles when designing, developing, integrating or using artificial intelligence (AI) systems to: ● achieve better outcomes ● reduce the risk of negative impact ● practice the highest standards of ethical business and good governance"
Bertelsmann Stiftung & iRights.Lab 2019: *Algo.Rules : Rules for the Design of Algorithmic Systems*	"Algorithmic systems are being implemented in a growing number of areas and are being used to make decisions that have a profound impact on our lives. They involve opportunities as well as risks. It is up to us to ensure that algorithmic systems are designed for the benefit of society. The individual and collective freedoms and rights that comprise human rights should be strengthened, not undermined, by algorithmic systems. Regulations designed to protect these norms must remain enforceable. To achieve this objective, we've developed the following Algo.Rules together with a variety of experts and the interested public [...] The Algo.Rules are a catalogue of formal criteria for enabling the socially beneficial design and oversight of algorithmic systems. They provide the basis for ethical considerations as well as the implementation and enforcement of legal frameworks"
China Ministry of Science and Technology expert committee 2019: *Governance*	"The global development of artificial intelligence (AI) has entered a new stage, presenting new features such as cross-domain integration, human-machine cooperation, and swarm integrated intelligence. It is profoundly changing the life of human society and changing the world. In order to promote the healthy development of a new generation of AI; better coordinate the relationship between development and governance,

principles for a New Generation of AI	ensure that AI is safe/secure, reliable, and controllable; promote economically, socially, and ecologically sustainable development; and jointly build a community of common destiny for humanity; various parties related to AI development should adhere to the following principles"
IBM 2019: *Everyday Ethics for Artificial Intelligence*	"As designers and developers of AI systems, it is an imperative to understand the ethical considerations of our work. A techcentric focus that solely revolves around improving the capabilities of an intelligent system doesn't sufficiently consider human needs. An ethical, human-centric AI must be designed and developed in a manner that is aligned with the values and ethical principles of a society or community it affects. Ethics is based on well-founded standards of right and wrong that prescribe what humans ought to do, usually in terms of rights, obligations, benefits to society, fairness, or specific virtues. To create and foster trust between humans and machines, you must understand the ethical resources and standards available for reference during the designing, building, and maintenance of AI" (p. 8)
IEEE: *A Call to Action for Businesses Using AI: Ethically Aligned Design for Business*	"[...] if you are working in the world of AI, you are, in fact, working for the future of humanity, so you need to embed ethics practices across all teams responsible for these types of systems. An ethical, human-centric AI must be designed and developed in a manner that is aligned with the values and ethical principles of the society or community it affects. To be human-centered, businesses must first establish a culture of trust, transparency, and accountability internally in order to effectively spread these values externally" (p. 3)
ITI (Information Technology Industry Council): *AI Policy Principles : Executive Summary*	"Highly autonomous AI systems must be designed consistent with international conventions that preserve human dignity, rights, and freedoms. As an industry, it is our responsibility to recognize potentials for use and misuse, the implications of such actions, and the responsibility and opportunity to take steps to avoid the reasonably predictable misuse of this technology by committing to ethics by design" (p. 1)
OECD 2015: *Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity*	"Digital security risk management should be implemented in a manner that is consistent with human rights and fundamental values recognised by democratic societies, including the freedom of expression, the free flow of information, the confidentiality of information and communication, the protection of privacy and personal data, openness and fair process. Digital security risk management should be based on ethical conduct which respects and recognises the legitimate interests of others and of the

	society as a whole. Organisations should have a general policy of transparency about their practices and procedures to manage digital security risk"
Global Alliance for Genomics and Health (GA4GH): *Framework for Responsible Sharing of Genomic and Health-Related Data*	"The purpose of this Framework is to provide a principled and practical framework for the responsible sharing of genomic and health-related data. Its primary goals are to: i. Protect and promote the welfare, rights, and interests of individuals from around the world in genomic and health-related data sharing, particularly those who contribute their data for biomedical research; ii. Complement laws and regulations on privacy and personal data protection, as well as policies and codes of conduct for the ethical governance of research; iii. Foster responsible data sharing and oversight of research data systems; iv. Establish a framework for greater international data sharing, collaboration and good governance; v. Serve as a dynamic instrument that can respond to future developments in the science, technology, and practices of genomic and health-related data sharing; vi. Serve as a tool for the evaluation of responsible research by research ethics committees and data access committees; and vii. Provide overarching principles to be respected in developing legally-binding tools such as data access agreements" (p. 2)
Kontekstspecifikke, implicitte definitioner, forskningslitteratur	
Heidelberg, Kelman, Hopkins, Allen 2020: *The evolution of data ethics in clinical research and drug development*	"Demystifying the use of emerging technologies and data sources, exploring ethical tensions, and considering the perspectives of patients, academics, industry researchers, and governments are starting points for building trust and evolving data ethics [...] Data ethics frameworks can provide guidelines for organizations to determine where, when, why, and how patient data will be used and protected, with clear benefits to society identified to help patients live longer, have better lives, and evolve the practice of medicine" (p. 2-5)
Salerno et al 2017: *Ethics, Big Data and Computing in Epidemiology and Public Health*	"Epidemiologists and research ethics board (REB)/institutional review board (IRB) members have a professional obligation to understand the ethical dimensions of the potential informational benefits and harm that big data present. We must prepare epidemiologists with the ethical tools they need to work in the new big data era. Explicit ethics training is no longer optional and should be part of every epidemiologist's training. In addition, REBs/IRBs should familiarize themselves with big data and the ethical dimensions of its use" (p. 298)

Berman, G & Albright, K. Unicef 2017: *Children and the Data Cycle: Rights and Ethics in a Big Data World*	"With this perspective in mind, fundamental questions need to be raised as to how best translate universal principles regarding the rights of the child and traditional ethical frameworks for offline data collection, analysis and regulation into an online environment. This includes ascertaining how to uphold such rights, and balancing the risks and opportunities for children that engagement may bring [...]" (p. 9)
E. Mandinach, B. Parton, E. Gummer, R. Anderson 2015: *Ethical and appropriate data use requires data literacy*	"Ethical and responsible data use is part of knowing how to use data, and that knowledge focuses on how to protect student privacy and maintain confidentiality of student data. Such knowledge includes how and when to discuss students' performance, behavior, attitudes, etc. with other teachers, administrators, and parents. It also includes knowing how to remove identifying information from a student record and how to maintain proper student records — whether electronic or in paper-and-pencil format. It includes knowing who has access to student records and when, how, and the process by which to release data or results. Responsible data use also includes knowing when and when not to discuss a student's performance in public" (p. 26)
J. Fairfield & H. Shtein 2014: *Big Data, Big Problems: Emerging Issues in the Ethics of Data Science and Journalism*	"Much of social science ethics focuses on rights and responsibilities toward the individual human participant. Big data as a technique does not accommodate this well. There can be millions of research subjects, yet none of them has given traditional informed consent. The traditional focus of social science has been on physical, rather than informational harms, and on not harming individuals, but big data impacts communities as much (or more) than individuals. Yet the notion that information is not a cognizable harm is not supportable in the context of an information-based society. This technological shift requires a rethinking of how ethical principles are carried out" (p. 39)
S. Leonelli 2016: *Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems*	"I propose a framework for the enforcement of ethical oversight over the dissemination and use of Big and Open Data, which is grounded on the importance of encouraging critical thinking and ethical reflection among the researchers involved in data processing practices, and aims to improve both the social impact and the scientific quality of data science practices and outputs [...] ethical reasoning should be an integral part of data science, which helps researchers to critically valuate and discuss the allocation of responsibilities and accountabilities within highly distributed and globalized trajectories of data production, dissemination and re-use" (p. 3)
C. Drew 2016:	"[...] Government Data Science Partnership has taken an open, evidence-

Data science ethics in government	based and user-centred approach to creating an ethical framework. It is a practical document that brings all the legal guidance together in one place, and is written in the context of new data science capabilities [...] a renewed deal between the citizen and state on data, to maintain and solidify trust in how we use people's data for social good [...] setting out clear guidance that brings together the relevant laws and best practice, gives data scientists and their teams robust principles to work with [...] The framework sets out how we need to balance public benefit (principle 1) with the risks to privacy (principle 2), validity and unintended consequences (principle 3)" (p. 1-7)
Xafis, Schaefer et al. 2019: *An Ethics Framework for Big Data in Health and Research*	"Ethical decision-making frameworks assist in identifying the issues at stake in a particular setting and thinking through, in a methodical manner, the ethical issues that require consideration as well as the values that need to be considered and promoted [...] This paper sets out an Ethics Framework for Big Data in Health and Research [...] It presents the aim and rationale for this framework supported by the underlying ethical concerns that relate to all health and research contexts. It also describes a set of substantive and procedural values that can be weighed up in addressing these concerns, and a step-by-step process for identifying, considering, and resolving the ethical issues arising from big data uses in health and research" (p. 227)

Bilag 1: Litteraturstudie 2 (Dataetiske værdier og principper)

Litteraturstudiet er foretaget ved gennemsyn af kilder fra litteraturstudie 1 (Se Dataetisk Råds analyse "Hvad er dataetik?") og ved nye søgninger på scholar.google.com, google.com og forskningsdatabasen.dk i oktober 2020.

Litteraturstudiet benyttede følgende søgeord på hver af de tre databaser:

- "Data ethics guideline"
- Data ethics guidelines
- "data ethics principles"
- "Data ethics framework"
- Data ethics statement
- "data ethics values"
- "Data ethics guide"
- "Dataetik guideline"
- Dataetik retningslinjer
- "Dataetiske retningslinjer"
- Dataetik principper
- Dataetik værdier
- "Dataetiske principper"
- "Dataetiske værdier"
- AI ethics guide
- AI ethics principles
- AI ethics values
- AI ethics guideline
- AI ethics framework
- AI statement
- AI etik principper
- AI etik værdier
- AI etik retningslinjer

Søgningerne genererede forventeligt mange hundrede resultater. Disse resultater er filtreret to gange. Ved de umiddelbare søgninger frasorteredes indlysende irrelevante resultater. De tilbageværende resultater blev indhentet og studeret. Ved læsning af disse resultater blev yderligere en håndfuld resultater sorteret fra som irrelevante.

Litteraturstudiet har identificeret 268 eksempler på behandling af dataetiske værdier og principper i relevante tekster. Heraf 43 i akademiske kilder, og 225 i policy-dokumenter og øvrige kilder. Teksternes behandling af værdier og principper er nedenfor angivet i citat, med henvisning til kildens forfatter, årstal, titel og hyperlink..

Kilderne er sorteret efter tema. Disse temaer er baseret på en klyngeanalyse, i hvilken alle kilder først blev gennemset med henblik på at markere de enkelte kilder med nøgleord. Dernæst blev en

finsortering foretaget med inddeling af kilderne i de 13 etiske værdier og principper, som er behandlet i denne analyse:

- Ansvarlighed
- Autonomi
- Dataetisk bevidsthed
- Demokrati
- Fairness
- Gennemsigtighed
- Godgørenhed
- Lighed
- Privatliv
- Retfærdighed
- Sikkerhed
- Velfærd
- Værdighed

Kilde	Citat
Ansvarlighed	
Accenture 2016: Universal principles of data ethics : 12 guidelines for developing ethics codes	"There is no such thing as raw data—all datasets and accompanying analytic tools carry a history of human decision-making. As much as possible, that history should be auditable, including mechanisms for tracking the context of collection, methods of consent, the chain of responsibility, and assessments of quality and accuracy of the data." (s.8) "Data scientists and practitioners should accurately represent their qualifications, limits to their expertise, adhere to professional standards, and strive for peer accountability [...] The long-term success of the field depends on public and client trust. Data professionals should develop practices for holding themselves and peers accountable to shared standards." (s.9) "Aspire to design practices that incorporate transparency, configurability, accountability, and auditability [...] Not all ethical dilemmas have design solutions, but being aware of design practices can break down many of the practical barriers that stand in the way of shared, robust ethical standards. Data ethics is an engineering challenge worthy of the best minds in the field." (s.9)
ACM US Public Policy Council 2017: Statement on Algorithmic Transparency and Accountability	"Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected." "Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results."
Amsterdam Economic Board (TADA): The 6 principles of our manifesto	"Citizens and users have control over the design of our digital city. The government, civil society organizations and companies facilitate this. They monitor the development process and the resulting social consequences."
Australian Government: AI Ethics Principles	"Those responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled." "This principle aims to acknowledge the relevant organisations' and individuals' responsibility for the outcomes of the AI systems that they design, develop, deploy and operate [...] This includes both before and after their design, development, deployment and operation. The organisation and individual accountable for the decision should be identifiable as necessary. They must consider the appropriate level of human control or oversight for the particular AI system or use case. AI systems that have a significant impact on an individual's rights should be accountable to external review, this includes providing timely, accurate, and complete information for the purposes of independent oversight bodies."
Bertelsmann Stiftung & iRights.Lab 2019: Algo.Rules :	"A natural or legal person must always be held responsible for the effects involved with the use of an algorithmic system." (s.4)

Design rules for algorithmic systems	"Accountability must be clearly assigned. The accountable person must be aware of the responsibilities associated with their tasks. This also applies to responsibilities that are shared by several people or organizations. The allocation of responsibility must be fully documented and transparent for internal and external parties. Responsibility may not be transferred to the algorithmic system itself, users or people who are affected by the system." (ibid.) "The effects of an algorithmic system must be reviewed on a regular basis." (s.6) "An algorithmic system must be subject to active monitoring in order to determine whether the targeted objectives are actually achieved, and the use of the system does not violate existing legislation. Taking the appropriate technological precautions, external bodies should be able to conduct an independent, comprehensive and effective audit of an algorithmic system without compromising legitimate concerns regarding business confidentiality. Should a negative impact be determined, the cause must be identified and the algorithmic system adapted accordingly." (ibid.) "Establish complaint mechanisms [...] If an algorithmic system results in a questionable decision or a decision that affects an individual's rights, it must be possible to request an explanation and file a complaint." (s.6) "The person or organization using an algorithmic system must provide an easily accessible means of contact. First, those affected must be able to request appropriate and detailed information regarding a specific decision and the considerations that have fed into it. This should be an option also for organizations acting in their legitimate interest and for situations in which an organization acts on the behalf of an individual. Second, there must be an easily accessible and effective way to lodge a complaint. Complaints and actions taken must also be documented." (ibid.)
Chinese Expert Group 2019: Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence	"AI developers, users, and other interested parties should possess a strong sense of social responsibility and self-discipline, and strictly abide by laws, regulations, ethics, morals, standards, and norms. Establish an AI accountability mechanism to clarify the responsibilities of developers, users, beneficiaries, etc. The AI application process should ensure the human right to know and give notice of possible risks and impacts. Prevent the use of AI for illegal activities."
Danish Expert Group on Data ethics 2018: Data for the Benefit of the People : Recommendations from the Danish Expert Group on Data Ethics / Data i menneskets tjeneste : Anbefalinger fra Ekspertgruppen om dataetik	"All sides must be responsible for the consequences of their technological solutions." (s.8) / "Alle led skal være ansvarlige for konsekvenserne af deres teknologiske løsninger." (s.8)
Den danske regering 2019: Dansk National Strategi for Kunstig Intelligens	"Alle led skal være ansvarlige for konsekvenserne af deres udvikling og anvendelse af kunstig intelligens, dvs. blandt andet udviklere, samarbejdspartnere, anvendere, myndigheder og virksomheder. Det skal ved beslutninger og beslutningsunderstøttelse truffet af kunstig intelligens være muligt at stille mennesker til ansvar." (s.28)

FAT/ML: Principles for Accountable Algorithms and a Social Impact Statement for Algorithms	“Make available externally visible avenues of redress for adverse individual or societal effects of an algorithmic decision system, and designate an internal role for the person who is responsible for the timely remedy of such issues.”
Future of Life 2017: ASILOMAR AI PRINCIPLES	“Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.” “Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.”
Google 2018: Objectives for AI Applications	"We will design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal. Our AI technologies will be subject to appropriate human direction and control."
Government of Canada: Our guiding principles	"Provide sufficient training so that government employees developing and using AI solutions have the responsible design, function, and implementation skills needed to make AI-based public services better."
Government of UK 2019: Understanding artificial intelligence ethics and safety	"It is important that there is clear responsibility for and ownership of AIDA-driven decisions within an AIDA firm, with appropriate internal approving authorities for the use of AIDA. Such accountability applies to all uses of AIDA, whether internally developed or externally sourced."
IBM 2019: Everyday Ethics for Artificial Intelligence	"AI designers and developers are responsible for considering AI design, development, decision processes, and outcomes." (s.14) "Human judgment plays a role throughout a seemingly objective system of logical decisions. It is humans who write algorithms, who define success or failure, who make decisions about the uses of systems and who may be affected by a system’s outcomes. Every person involved in the creation of AI at any step is accountable for considering the system’s impact in the world, as are the companies invested in its development." (ibid.) Til princippet hører også "Recommended actions to take" og eksempler (se s. 16-18).
ITI (Information Technology Industry Council): AI Policy Principles : Executive Summary	"The use of AI to make autonomous consequential decisions about people, informed by – but often replacing decisions made by – human-driven bureaucratic processes, has led to concerns about liability. Acknowledging existing legal and regulatory frameworks, we are committed to partnering with relevant stakeholders to inform a reasonable accountability framework for all entities in the context of autonomous systems." (s.4)
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	“Accountability By Design: All AI systems must be designed to facilitate end-to-end answerability and auditability. This requires both responsible humans-in-the-loop across the entire design and implementation chain and activity monitoring protocols that enable end-to-end oversight and review.” (s.12)

Microsoft: Microsoft AI Principles	“There should be a common set of standards in which companies are held accountable for the use and impact of their AI technology.”
Mishra 2020: International Trade Law Meets Data Ethics: A Brave New World	"The principle of algorithmic accountability is highly debated. The basis of algorithmic accountability is that data-driven technologies the service or technology suppliers should be able to explain how their algorithms and technical designs use and process data to generate certain results, and, if and when needed, rectifiable to ensure compliance with laws and regulations. This kind of transparency and explainability can in turn help check instances of unfair or discriminatory outcomes affecting specific individuals (for example, if they violate human rights or constitutional rights), providing redress to the affected individuals, and rectification of malfunctioning software and other digital services. As the use of AI is increasingly common for various public and commercial purposes, the principle of algorithmic accountability plays a significant role in addressing ‘blackbox malfunctioning’ i.e. where the lack of transparency regarding the functioning of AI algorithms leads to unfair or discriminatory decisions/outcomes. Thus, the principle of algorithmic accountability can also be viewed from the perspective of preserving international or domestic human rights." (s.12)
Montreal Declaration Responsible AI: THE DECLARATION	“Only human being can be held responsible for decisions stemming from recommendations made by AI, and the actions that proceed therefrom. In all areas where a decision that affects a person's life, quality of life, or reputation must be made, where time and circumstance permit, the final decision must be taken by a human being and that decision should be free and informed. The decision to kill must always be made by human beings, and responsibility for this decision must not be transferred to an AI. People who authorize AI to commit a crime or an offence, or demonstrate negligence by allowing Ai to commit them, are responsible for this crime or offence. When damage or harm has been inflicted by an AI, and the AI is proven to be reliable and to have been used as intended, it is not reasonable to place blame on the people involved in its development or use.”
Norwegian Ministry of Local Government and Modernisation 2020: National Strategy for Artificial Intelligence	“AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.” “Individuals must have the right not to be subject to automated processing when the decision made by the system significantly affects them. Individuals must be included in decision-making processes to assure quality and give feedback at all stages in the proces ('human-in-the-loop').”
OECD 2019: Recommendation of the Council on Artificial Intelligence	“An AI system should be deployed only after an adequate evaluation of its purpose and objectives, its benefits, as well as its risks. Institutions must be responsible for decisions made by an AI system.”
Regulatory and Ethics Working Group of the Global Alliance for Genomics and Health 2014: Framework for Responsible Sharing of	"• Put in place systems for data sharing that respect this Framework. • Track the chain of data access and/or exchange to its source. • Develop processes to identify and manage conflicts of interest. • Implement mechanisms for handling complaints related to data misuse; for identifying, reporting and managing breaches; and for instituting appropriate sanctions." (s.4)

Genomic and Health-Related Data	
The Linux Foundation: Data Values and Principles	"Behave ethically and transparently, fix mistakes quickly, and hold ourselves and others accountable."
The Public Voice 2018: Universal Guidelines for Artificial Intelligence	“Institutions must be responsible for decisions made by an AI system.” “The institution responsible for an AI system must be made known to the public.” “All individuals have the right to a final determination made by a person.”
The Task Force on Artificial Intelligence of the Agency for Digital Italy 2018: White Paper on Artificial Intelligence at the service of citizens	"The examples just mentioned highlight the strong impact that Artificial Intelligence has on the decision-making activity of public entities. Both when it acts as an assistant to human beings as well as an autonomous entity, AI generates effects on the lives of people in relation to which it is necessary to be able to establish legal liability. Nevertheless, the ownership of the latter is not clearly identifiable, since it could be attributed to the producer or to the owner of the Artificial Intelligence, or even to its end user. Those who design AI systems can be responsible for design or implementation defects, but not for behaviour caused by inadequate instruction datasets. Can a public decision-maker be considered politically responsible for the decisions made on the basis of algorithms that process data affected by the bias mentioned above? What type of responsibility can there be for Public Administration? If a robot hurts someone, who should be held responsible and who, if anyone, has the obligation to compensate the victim (and with which assets)? Can the public decision-maker transfer his political responsibility to an AI system that does not respond to a clear principle of representation? Is it ethically sustainable that, in order to improve the efficiency and effectiveness of measures, certain important choices can be made with the influence of an AI or even completely delegating them to the AI? And in trusting an AI system, how can its consistency be controlled over time? These are just some of the issues that emerge in this area and highlight the need to establish some principles for the use of AI technologies in a public context." (s.36)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	“Arrangements should be developed that will make possible to attribute accountability for AI-driven decisions and the behaviour of AI systems.”
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	"Accountability refers to the ability to scrutinise judgements, decisions and actions, and for decision-makers to be held responsible for their consequences." (s.246)
Autonomi	
Accenture 2016: Universal principles of data ethics : 12 guidelines for developing ethics codes	"Data professionals should strive to use data in ways that are consistent with the intentions and understanding of the disclosing party. Many regulations govern datasets on the basis of the status of the data, such as “public,” “private” or “proprietary.” However, what is done with datasets is ultimately more consequential to subjects/users than the type of data or the context in which it is collected. Correlative uses of repurposed data in research and

	industry represents both the greatest promise and the greatest risk posed by data analytics." (s.8)
Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	"[...] traditional modes of ensuring consent and safeguarding child rights are neither possible nor feasible in an online environment. Under many national and international legislative and regulatory frameworks, guardians or parents are responsible for providing parental consent for the collection of data from children under eighteen or the relevant age of majority [...] informed consent or assent (in cases where children are not legally permitted to provide informed consent) should be received from the child, following clear articulation and full disclosure of the planned use of the data collected, communicated using language and methods that children can easily understand. Essentially, clear guidance must be provided to children to enable them to withdraw from participation or refuse to provide data at any point in the process." (s.12)
Brakewood & Poldrack 2013: The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data	“The Belmont report states that respect for persons is a principle that has two “ethical convictions”: people should be able to make autonomous decisions and people with limited autonomy should be protected. Others have argued that a broader application of the principle of respect for persons should also include recognizing the inherent dignity of people, recognizing that autonomy is only one aspect of the principle. Autonomous decision-making means that a subject needs to have the ability to actually act on that decision. Subjects must have sufficient information in a format they understand, and they must have the ability to choose.” (s.673)
Danish Expert Group on Data ethics 2018: Data for the Benefit of the People : Recommendations from the Danish Expert Group on Data Ethics / Data i menneskets tjeneste : Anbefalinger fra Ekspertgruppen om dataetik	"People must retain the most control possible over their own data." (s.8) / "Mennesket skal bevare mest mulig kontrol over egne data." (s.8)
Datenethik Kommission 2019: Opinion of the Data Ethics Commission	"The opportunity for self-determination is inextricably linked with human dignity. Humans express their freedom by determining their life goals and the way they lead these lives, as a basis for determining, developing and enacting the very essence of their self. A society that takes freedom seriously must put in place a framework within which its citizens can develop freely and respect each other’s freedom, despite all their differences. For example, if people are to lead a self-determined life and develop in freedom, technical systems must not restrict and control human avenues for action without an ethically meaningful reason. Self-determination must not be viewed solely through an individualistic lens – humans are relational beings whose life unfolds through social interactions with others, on the basis of manifold reciprocal links and influences." (s.43)

Den danske regering 2019: Dansk National Strategi for Kunstig Intelligens	"Menneskets autonomi prioriteres i udvikling og anvendelse af kunstig intelligens. Mennesket skal som i dag kunne træffe oplyste og selvstændige valg, uden at kunstig intelligens fjerner menneskets selvbestemmelse." (s.28)
Montreal Declaration Responsible AI : THE DECLARATION	"AI development and use must not lead to the homogenization of society through the standardization of behaviours and opinions. From the moment algorithms are conceived, AI development and deployment must take into consideration the multitude of expressions of social and cultural diversity present in the society. AI must avoid using acquired data to lock individuals into a user profile, fix their personal identity, or confine them to a filtering bubble, which would restrict and confine their possibilities for personal development - especially in fields such as education, justice, or business."
EDPS Ethics Advisory Group 2018: Towards a digital ethics	"Like dignity, freedom is one of the foundational values of the European Union and a pillar of the common provisions of the Treaty of European Union. Since 1999, it has, in addition, been the core feature of the Schengen political program of offering all EU citizens an 'area of freedom, security and justice'. In its core European ethical and judicial formulations, freedom is understood as a determining, positive right. It determines in the sense that it is not regarded as unconditional, but rather made meaningful by its insertion into the European system of values, underpinning a specific system of laws, directives, communication and international obligations" (s.17)
EU Commision for the Efficiency of Justice 2018: European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment	"Preclude a prescriptive approach and ensure that users are informed actors and in control of the choices made ■User autonomy must be increased and not restricted through the use of artificial intelligence tools and services. ■Professionals in the justice system should, at any moment, be able to review judicial decisions and the data used to produce a result and continue not to be necessarily bound by it in the light of the specific features of that particular case. ■The user must be informed in clear and understandable language whether or not the solutions offered by the artificial intelligence tools are binding, of the different options available, and that s/he has the right to legal advice and the right to access a court. S/he must also be clearly informed of any prior processing of a case by artificial intelligence before or during a judicial process and have the right to object, so that his/her case can be heard directly by a court within the meaning of Article 6 of the ECHR. ■Generally speaking, when any artificial intelligence-based information system is implemented there should be computer literacy programmes for users and debates involving professionals from the justice system." (p- 12)
Fairfield & Shtein 2014: Big Data, Big Problems : Emerging Issues in the Ethics of Data Science and Journalism	"The autonomy principle denotes respect for persons. Individuals are to be treated as autonomous agents. In practice, the autonomy principle translates to the need to obtain informed consent." (s.40)
Forsikring & Pension 2019: Databrug og dataetik : Dilemmaer og mulige	"Ændrer jeg adfærd til det bedre eller værre (afhængigt af, hvor fornuftig adfærdsændringen er) som følge af, at jeg deler data og bliver overvåget? Og er jeg tilskyndet økonomisk, eller via en form for værditransaktion, til at dele

positioner for forsikrings- og pensionsbranchen	mere data, tilbageholde dem eller manipulere mine data til fordel eller ulempe for mig selv og/eller andre?”
Future of Life 2017: ASILOMAR AI PRINCIPLES	“People should have the right to access, manage and control the data they generate, given Ai systems' power to analyze and utilize that data.” “The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.”
Government of the Netherlands 2019: Strategic Action Plan for Artificial Intelligence	"This includes the risk of dehumanisation and the influence of AI on making choices" (s.41)
Herschel & Miori 2017: Ethics & Big Data	"Kantian analysis argues that one should always respect the autonomy of other people, treating them as ends in themselves and never only as means to an end. With Big Data, this would be a difficult case to make. Since data is routinely collected and analyzed to assess individuals without their consent, organizations employing Big Data are not respecting the autonomy of people and they are in fact using personal data as a means to an end to further the organization's self-interest. The nature of Big Data is that in general, people typically do not opt-in to their data collection and exploitation, demonstrating their consent and hence shared responsibility. This means that by default, their privacy is compromised for the gain of another. Organizations utilizing Big Data may argue that they post information online informing consumers that the data they capture from a user's online behavior patterns may be used to offer new products or services. Some even provide their customers the option to opt-out of the firm's ability to share their information with organization's business partners. The fact is, however, that practically speaking, no one has the ability to determine how their data is actually shared and used, because the Big Data space is too big and there is no mechanism affording the individual the ability to actively monitor and control their private information. Hence to a great extent, individuals are blind to the sharing of their digitized data. While individuals may employ digital services to warn them of identity theft, to monitor credit issues, or to inform them when they are mentioned in postings, they are typically forced to be reactive rather than proactive posture in responding to information that may affect their privacy and security. Big Data compromises the old adage that state “Treat people how you want to be treated”. This phrase speaks to the Kantian notion that one should act only on the moral rules that you can imagine everyone else following. However, with Big Data individuals are frequently represented simply as data points that are then used to manipulate what the person will view in the future. That is, information is presented to individuals online that Big Data calculations determine best reflects their projected preferences based upon their previous search and online page view history. This algorithmic manipulation presumes the will of the individual without their explicit consent. Using a Kantian viewpoint, one might ask whether everyone should assent to a rule that states that everyone's information can be shared with or without their permission, regardless whether it is accurate or inaccurate, complete or incomplete, current or dated, and this information can be used to influence and represent peoples' behavior and interests with or without their consent. This is probably unlikely, otherwise there would not be so many concerns expressed about Big Data and privacy rights and protection. The point here is that Kantianism

	provides a relatively straightforward means for discussing the ethics of Big Data. It asserts that all people are rational, autonomous beings having moral worth and everyone is held to the same universal moral guideless. Because of this, Big Data is problematic for Kantian belief's because the actions associated with Big Data challenge the rights and fair treatment of the individual." (s.34)
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	"• Ensure their abilities to make free and informed decisions about their own lives • Safeguard their autonomy, their power to express themselves, and their right to be heard • Secure their capacities to make well-considered and independent contributions to the life of the community • Support their abilities to flourish, to fully develop themselves, and to pursue their passions and talents according to their own freely determined life plans." (s.10)
Malta AI Taskforce 2019: Towards Trustworthy AI - Malta Ethical AI Framework for public consultation	"Humans interacting with AI systems must be able to keep full and effective self-determination over themselves."
Medium / Towards data science 2018: 5 Principles for Big Data Ethics	"Big data analytics can moderate and even determine who we are before we make up our own minds. Companies need to begin to think about the kind of predictions and inferences that should be allowed and the ones that should not."
Mittelstadt & Floridi 2016: The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts	"Half of the literature addresses issues of informed consent. The concept does not cleanly transfer to research involving Big Data for a variety of reasons. Historically, consent is taken for participation in a single study, not covering unrelated investigations resulting from sharing, aggregating, or even repurposing data within the wider research community. This form of consent is problematic because Big Data is intended by design to reveal unforeseen connections between data points. This means that both what the data reveals about the subject and its utility in future research present greater uncertainty than normal at the time of consent [...] 'consent' cannot be 'informed' in the sense that data subjects cannot be told about future uses and consequences of their data, which are unknowable at the time the data is collected or aggregated [...] Such barriers often lead to biobanks employing a broad type of consent covering all future research activities. However, this approach has been recognised to limit the autonomy of data subjects. Tiered consent can also be used, which provides the data subjects with options for permitting specific uses of their data—for example to allow the data to be used in cancer research but not in genomic research—or to require specific re-consent for future uses rather than blanket consent for all potential uses."
Montreal Declaration Responsible AI : THE DECLARATION	"AI must allow individuals to fulfill their own moral objectives and their conception of a life worth living. AI must not be developed or used to impose a particular lifestyle on individuals, whether directly or indirectly, by implementing oppressive surveillance and evaluation or incentive mechanisms. Public institutions must not use AI to promote or discredit a particular conception of the good life. It is crucial to empower citizens regarding digital

	technologies by ensuring access to the relevant forms of knowledge, promoting the learning of fundamental skills (digital and media literacy), and fostering the development of critical thinking. AI must not be developed to spread untrustworthy information, lies, or proaganda, and should be designed with a view to containing their dissemination. The development of AI must avoid creating dependencies through attention-capturing techniques or the imitation of human characteristics (appearance, voice, etc) in ways that could cause confusion between AI and humans."
Richards & King 2014: Big Data Ethics	"Whereas privacy harkens from the right to be let alone, identity hails from the fundamental right to define who we are [...] privacy protections are not enough in our new age of the big metadata computer because big data analytics can compromise identity by allowing institutional surveillance to moderate and even determine who we are before we make up our own minds [...] we must begin to think about the kinds of big data predictions and inferences that we will allow and the ones that we should not." (s.396)
Slade & Prinsloo 2013: Learning analytics: ethical issues and dilemmas	"In stark contrast to seeing students as producers and sources of data, learning analytics should engage students as collaborators and not as mere recipients of interventions and services. Not only should students provide informed consent regarding the collection, use and storage of data, but they should also voluntarily collaborate in providing data and access to data to allow learning analytics to serve their learning and development, and not just the efficiency of institutional profiling and interventions. […] To value students as agents, making choices and collaborating with the institution in constructing their identities (however transient) can furthermore be a useful (and powerful) antidote to the commercialization of higher education in the context of the impact of skewed power relations, monitoring and surveillance." (s.12-13) "Integral in learning analytics is the notion of student identity. It is crucial to see student identity as a combination of permanent and dynamic attributes. During students' enrollment, their identities are in continuous flux and as such, they find themselves in a "Third Space" where their identities and competencies are in a permanent liminal state. The ethical implications of this are that learning analytics provides a snapshot view of a learner at a particular time and context. This not only necessitates the need for longitudinal data but has implications for the storage and permanency of data. Mayer-Schönberger warns that forgetting is a "fundamental human capacity." Students should be allowed to evolve and adjust and learn from past experiences without those experiences, due to their digital nature, becoming permanent blemishes on their development history. […] Data collected through learning analytics should therefore have an agreed-upon lifespan and expiry date, as well as mechanisms for students to request data deletion under agreed-upon criteria." (s.13)
The Linux Foundation : Data Values and Principles	"Present our work in ways that empower others to make better-informed decisions."
Tranberg, Hasselbalch, Olsen & Byrne / DataEthics.eu - The independent Thinkdotank 2018: DATAETHICS –	"Humans should be in control of their data and empowered by their data. A person's self-determination should be prioritised in all data processes and the person should be actively involved in regards to the data recorded about them.

Principles and Guidelines for Companies, Authorities & Organisation	The individual has the primary control over the usage of their data, the context in which his/her data is processed and how it is activated." (s.10)
UK Statistics Authority: Data Ethics	"The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately."
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	"AI should respect human autonomy by requiring human control at all times." "Algorithm awareness and a basic understanding of the workings of AI are needed to empower citizens."
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	"Liberty and autonomy are very closely related concepts. For the purpose of this document, we define liberty as the state of not being coerced by physical, legal, or social pressure into action by some outside influence. Autonomy is defined as the capacity of a person or group to be self-determining." (s.245) "Engagement is the meaningful involvement of stakeholders in the design and conduct of the data activities. Engagement goes beyond the dissemination of information and requires that data activities have been influenced in some way by the views of stakeholders." (s.246)
Zimmer 2018: Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity	"One of the foundations of research ethics is the idea of informed consent. Simply put, informed consent means that participants are voluntarily participating in the research with full knowledge of relevant risks and benefits. Providing informed consent typically includes the researcher proactively explaining the purpose of the research, the methods used, the possible outcomes of the research, as well as associated risks or harms that the participants might face. The process involves providing the subject clear and understandable explanations of these issues, providing sufficient opportunity to consider them before granting consent, and ensuring the subject has not been coerced into participating. Importantly, obtaining informed consent requires a verification of understanding and, thus, necessitates an ongoing communicative relationship between researchers and their participants. Obtaining consent in traditional research settings is typically done through a direct interaction between the researcher and the subject, through face-to-face mode, through telephone or video-conference scripts, or through mailed documents. The rise in Internet-based research—where researchers often interact with subjects asynchronously through online surveys or scrape data from subjects' social networking profiles—has introduced various challenges to the traditional approach to obtaining informed consent, including verifying the identity and demographic profile of subjects, ensuring comprehension of the consent form, and obtaining appropriate documentation of the consent. Various approaches and standards have emerged in response to these new challenges to obtaining informed consent in online environments, including providing a consent form prior to completing an online survey and requiring a subject to click "I agree" to proceed to the questionnaire, embedding implicit consent to research activities within other terms of use within a particular online service or platform, or deciding (rightfully or not) that some forms of online research are exempt from the need for obtaining informed consent." (s.3)

Demokrati	
DANSK IT's arbejdsgruppe for dataetik 2018: Dataetik : 18 dataetiske anbefalinger fra DANSK IT	"Tillid går begge veje, hvorfor der ved indførelse af sådanne mekanismer [fx kunstig intelligens i politiets forebyggende arbejde] tilsvarende bør indføres demokratiske processer og effektiv demokratisk kontrol."
Datenethik Kommission 2019: Opinion of the Data Ethics Commission	"Digital technologies are in a complex manner systemically relevant for the development of fundamental rights (in particular freedom of expression and information, (informational) self-determination, confidentiality of telecommunications, freedom of assembly and association, freedom of occupation and right to property), for democracy, for the safeguarding of diversity, for an open societal debate and for free and equal elections. For example, social media sites serve as a low-threshold opportunity for every citizen to participate in debates on the shape of our future, and as such should in principle be welcomed. At the same time, however, there is a risk that they may be used for manipulation and radicalisation. The State should take decisive action to counter these risks by adopting rules and setting up institutions capable of preventing undesirable developments and misuse." (s.46)
Future of Life 2017: ASILOMAR AI PRINCIPLES	"The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends."
Government of the Netherlands 2019: Strategic Action Plan for Artificial Intelligence	"In the application of AI, freedom of speech may come under pressure. This concerns access to information (e.g. personalisation and ordering of search results) on the one hand, and the operation of algorithms that automatically remove content on the other." (s.41)
Montreal Declaration Responsible AI : THE DECLARATION	"AI must not be developed or used with the aim of limiting the free expression of ideas or the opportunity to hear diverse opinions, both of which being essential conditions of a democratic society."
Norwegian Ministry of Local Government and Modernisation 2020: National Strategy for Artificial Intelligence	"The development and use of AI must foster a democratic and fair society by strengthening and promoting the fundamental freedoms and rights of the individual."
OECD 2019: Recommendation of the Council on Artificial Intelligence	"AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights. To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art."

UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	“AI should be developed, implemented and used in line with democratic principles.”
Etik	
Accenture 2016: Universal principles of data ethics : 12 guidelines for developing ethics codes	"Products and research practices should be subject to internal, and potentially external ethical review [...] Organizations should prioritize establishing consistent, efficient, and actionable ethics review practices for new products, services, and research programs. Internal peer-review practices can mitigate risk, and an external review board can contribute significantly to public trust." (s.9) “Always follow the law, but understand that the law is often a minimum bar [...] As digital transformations have become a standard evolutionary path for businesses, governments and laws have largely failed to keep up with the pace of digital innovation and existing regulations are often mis-calibrated to present risks. In this context, compliance means complacency. To excel in data ethics, leaders must define their own compliance frameworks that outperform legislated requirements." (s.8)
ACM US Public Policy Council 2017: Statement on Algorithmic Transparency and Accountability	"Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society."
Australian Government : AI Ethics Principles	"Throughout their lifecycle, AI systems should respect human rights, diversity, and the autonomy of individuals." "This principle aims to ensure that AI systems are aligned with human values. Machines should serve humans, and not the other way around. AI systems should enable an equitable and democratic society by respecting, protecting and promoting human rights, enabling diversity, respecting human freedom and the autonomy of individuals, and protecting the environment. Human rights risks need to be carefully considered, as AI systems can equally enable and hamper such fundamental rights. It’s permissible to interfere with certain human rights where it’s reasonable, necessary and proportionate."
Chinese Expert Group 2019: Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence	"AI development should begin from the objective of enhancing the common well-being of humanity; it should conform to human values, ethics, and morality, promote human-machine harmony, and serve the progress of human civilization; it should be based on the premise of safeguarding societal security and respecting human rights, avoid misuse, and prohibit abuse and malicious application."
Danish Expert Group on Data ethics 2018: Data for the Benefit of the People : Recommendations from the Danish Expert Group on Data Ethics /	"Societal progress in using data can be achieved through data ethical situations." (s.8) / "De samfundsmæssige fremskridt ved brug af data kan opnås ved brug af dataetiske løsninger." (s.8)

Data i menneskets tjeneste : Anbefalinger fra Ekspertgruppen om dataetik	
Den danske regering 2019: Dansk National Strategi for Kunstig Intelligens	"Kunstig intelligens kan være med til at skabe store fremskridt for samfundet. Der bør skabes tekniske og organisatoriske løsninger, der understøtter etisk ansvarlig udvikling og anvendelse af kunstig intelligens for at opnå størst mulig fremskridt for samfundet fx ved at bidrage til bedre service fra det offentlige og vækst i erhvervslivet." (s.29)
Drew 2016: Data science ethics in government	"Both the law and ethical practice require us to understand public perception so we can work out what we should do." (s.7)
EDPS Ethics Advisory Group 2018: Towards a digital ethics	"Personal or anonymous data are the new co-ordinates of social modelling. Big data rather than institutional or deliberative processes threaten to become the basis on which individuals are classified, evaluated, rewarded or punished. These same categories are used to evaluate the merits and needs of individuals or the opportunities or dangers underlying the lives they lead. In this view of 'data-driven governance', the question arises whether the individual human person as a legal subject has a future and how one can ensure that individuals are not viewed only as temporary data aggregates exploitable on an industrial scale rather than subjects in their own right. Interactions based on algorithmic profiling may exacerbate information imbalances between decision-making governments and companies on the one hand and individuals on the other hand. As a result, 'data-rich' public and private organisations will have greater ethical responsibilities towards citizens and customers. Digital ethics must identify new perspectives, potential and boundaries for dealing with data ethically, by formulating the terms of a proactive approach to ethics, beyond mere legal avoidance measures. As such it will set out the terms of a social innovation that parallels the rapid technological innovation we are experiencing on a daily basis." (s.19)
EU Commision for the Efficiency of Justice 2018: European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment	"Ensure that the design and implementation of AI tools and services are compatible with fundamental rights. ■The processing of judicial decisions and data must serve clear purposes, in full compliance with the fundamental rights guaranteed by the European Convention on Human Rights (ECHR) and the Convention on the Protection of Personal Data… ■When artificial intelligence tools are used to resolve a dispute or as a tool to assist in judicial decision-making or to give guidance to the public, it is essential to ensure that they do not undermine the guarantees of the right of access to the judge and the right to a fair trial (equality of arms and respect for the adversarial process). ■They should also be used with due respect for the principles of the rule of law and judges' independence in their decision-making process. ■Preference should therefore be given to ethical-by-design or humanrights-by-design approaches. This means that right from the design and learning phases, rules prohibiting direct or indirect violations of the fundamental values protected by the conventions are fully integrated." (s.8)

Future of Life 2017: ASILOMAR AI PRINCIPLES	“AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.”
IBM 2019: Everyday Ethics for Artificial Intelligence	"AI should be designed to align with the norms and values of your user group in mind." (s.20) "AI works alongside diverse, human interests. People make decisions based on any number of contextual factors, including their experiences, memories, upbringing, and cultural norms. These factors allow us to have a fundamental understanding of “right and wrong” in a wide range of contexts, at home, in the office, or elsewhere. This is second nature for humans, as we have a wealth of experiences to draw upon. Today’s AI systems do not have these types of experiences to draw upon, so it is the job of designers and developers to collaborate with each other in order to ensure consideration of existing values. Care is required to ensure sensitivity to a wide range of cultural norms and values. As daunting as it may seem to take value systems into account, the common core of universal principles is that they are a cooperative phenomenon. Successful teams already understand that cooperation and collaboration leads to the best outcomes." (s.20) Til princippet hører også "Recommended actions to take" og eksempler (se s. 22-24).
IEEE : A Call to Action for Businesses Using AI : Ethically Aligned Design for Business	“Modify existing design and development practices to include ethical AI considerations: • Work with executives to identify ethical principles based on the company’s existing core values. Identify critical inflection points to see which principles can make the biggest impact when identified and addressed. Some teams may start with a list of commonly agreed-upon AI principles. • Ensure stakeholder alignment on the potential risks and benefits of AI implementation at the beginning of every design phase. Create a taxonomy of risks in order to appropriately prioritize and address those you have identified. • Look for the most relevant subject matter experts who can provide deep and nuanced recommendations for each new design opportunity.”
ITI (Information Technology Industry Council): AI Policy Principles : Executive Summary	"We recognize our responsibility to integrate principles into the design of AI technologies, beyond compliance with existing laws. While the potential benefits to people and society are amazing, AI researchers and subject matter experts, and stakeholders should and do spend a great deal of time working to ensure the responsible design and deployment of AI systems. Highly autonomous AI systems must be designed consistent with international conventions that preserve human dignity, rights and freedoms. As an industry, it is our responsibility to recognize potential for use and misuse, the implications of such actions, and the responsibility and opportunity to take steps to avoid the reasonably predictable misuse of this technology by committing to ethics by design." (s.3)
Mishra 2020: International Trade Law Meets Data Ethics: A Brave New World	"The most fundamental principle in data ethics is the protection of human rights. A growing consensus exists that the use, processing and sharing of data in different digital technologies must comply with the basic principles of human rights. The key essence of a human rights-centric approach is protecting the dignity and individual rights of individuals, thus requiring governments to respect, protect and fulfill human rights. In adopting a human rights-centric

	approach, governments must address whether AI/ML and other data-driven applications are compatible with human rights; protect individuals from harm arising from data-driven technologies; and ensure access to adequate forms of redress to individuals adversely affected by data-driven technologies, such as challenging any discrimination or surveillance arising from unfair data practices, whether by governments or private companies." (s.7) "The principle of ethical design directly flows from a human rights-centric approach in data governance. It essentially means that technical designs and standards underlying data-driven technologies must comply with basic human rights. Some regulators ensure ethical design by enforcing rules requiring digital service suppliers to adopt and incorporate, by default, technologies and corporate policies that ensure privacy of individuals and protect their data from unauthorised intrusions. The idea behind this approach is that technological solutions to privacy or security issues are sustainable and more pragmatic to implement. The principle of ethical design promotes 'technological due process', in that designers are required to verify and establish that data-driven technologies function as expected. Some experts also call this the 'responsibility-by-design approach'. Ethical design can also facilitate more meaningful human control over certain aspects of designing of data-driven technologies, making it easier for engineers to explain technical designs, whether ex-ante or post-facto." (s.11)
Mittelstadt & Floridi 2016: The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts	"[...] The aforementioned complexity leads to several related problems. One is a tendency, particularly in mass media and industry, to view Big Data as 'objective' or as revealing objective truths without the need for human interpretation. This 'mythological' view of Big Data as the 'end of theory' creates ethical concerns regarding justification of increasingly pervasive and unbounded secondary manipulation and aggregation of data when Big Data practices are seen as the future of science and scientific discoveries."
SAP AI Ethics Steering Committee 2018: SAP's Guiding Principles for Artificial Intelligence	"We strive to create AI software systems that are inclusive and that seek to empower and augment the talents of our diverse usership. By providing human-centered user experiences through augmented and intuitive technologies, we leverage AI to support people in maximizing their potential. To achieve this, we design our systems closely with users in a collaborative, multidisciplinary, and demographically diverse environment."
Singapore Personal Data Protection Commission 2020: MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK second edition	"AI solutions should be human-centric [...] As AI is used to amplify human capabilities, the protection of the interests of human beings, including their well-being and safety, should be the primary considerations in the design, development and deployment of AI."
Smart Dubai : ARTIFICIAL INTELLIGENCE PRINCIPLES & ETHICS	"AI should be beneficial to humans and aligned with human values, in both the long and short term."
The Japanese Society for Artificial Intelligence 2017: The Japanese Society for	"As specialists, members of the JSAI shall recognize the need for AI to be safe and acknowledge their responsibility in keeping AI under control. In the development and use of AI, members of the JSAI will always pay attention to

Artificial Intelligence Ethical Guidelines	safety, controllability, and required confidentiality while ensuring that users of AI are provided appropriate and sufficient information.”
The Linux Foundation : Data Values and Principles	"Consider carefully the ethical implications of choices we make when using data, and the impacts of our work on individuals and society."
Tranberg, Hasselbalch, Olsen & Byrne / DataEthics.eu - The independent Thinkdotank 2018: DATAETHICS – Principles and Guidelines for Companies, Authorities & Organisation	"Human interests always prevail for [sic] institutional and commercial interests. People are not computer processes or pieces of software, but unique with empathy, self- determination, unpredictability, intuition and creativity and therefore have a higher status than machines.” “The human being is at the centre and have [sic] the primary benefit of data processing." (s.9) "Accountability is an organisation’s reflective, reasonable and systematic use and protection of personal data. Accountability is an integral part of all aspects of data processing, and efforts are being made to reduce the risks for the individual and to mitigate social and ethical implications. Sustainable personal data processing is embedded throughout the organisation and ensures ethical accountability in the short, medium and long term. An organisation’s accountability should also apply to subcontractor’s and partners’ processing of data." (s.11)
UK Government, Department for Digital, Culture, Media & Sport 2018: Guidance : Data Ethics Framework	"Embed data use responsibly [...] It is essential that there is a plan to make sure insights from data are used responsibly. This means that both development and implementation teams understand how findings and data models should be used and monitored with a robust evaluation plan." (s.5)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	“Developers and companies should take into consideration ethics when developing autonomous intelligent system.” “AI should be developed and implemented in accordance with international human rights standards.”
Willis 2013: Ethics, Big Data, and Analytics: A Model for Application	""Act on the maxim that you wish to have become a universal law." A college's administration has a unique duty once its chosen algorithms have demonstrated their effectiveness in predicting outcomes and providing actionable knowledge. At this point, the administration has a relational duty to make decisions in a way that is devoid of personal motivations. In decision-making based on analysis of big data, college administrations must consider carefully the implications of universally applying what is "known" through algorithmic logic." (s.5)
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	"Reflexivity refers to the process of reflecting on and responding to the limitations and uncertainties embedded in knowledge, information, evidence, and data. This includes being alert to competing and conflicting personal, professional, and organisational interests and to the management of associated biases. Reflexive institutions revise or create new policies and systems that change institutional processes and prompt further reflection and response." (s.246)
Fairness	

Australian Government : AI Ethics Principles	"Throughout their lifecycle, AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups." "This principle aims to ensure that AI systems are fair and that they enable inclusion throughout their entire lifecycle. AI systems should be user-centric and designed in a way that allows all people interacting with it to access the related products or services. This includes both appropriate consultation with stakeholders, who may be affected by the AI system throughout its lifecycle, and ensuring people receive equitable access and treatment."
Binns, R. (2018): Fairness in Machine Learning: Lessons from Political Philosophy	"Machine learning allows us to predict and classify phenomena, by training models using labelled data from the real world. When consequential decisions are made about individuals on the basis of the outputs of such models, concerns about discrimination and fairness inevitably arise [...] One question which immediately arises in such an endeavour is the need for formalisation. "What does it mean for a machine learning model to be 'fair' or 'non-discriminatory', in terms which can be operationalised?" "Current approaches to fair machine learning are typically focused on interventions at the data preparation, model-learning or post-processing stages. This is understandable given the typical remit of data scientists who are intended to carry out these processes. However, there is a danger that this results in an approach which focuses on a narrow, static set of prescribed protected classes, derived from law and devoid of context, without considering why those classes are protected and how they relate to the particular justice aspects of the application in question. Philosophical accounts of discrimination and fairness prompt reflection on these more fundamental questions, and suggest avenues for further consideration of what might be relevant and why. This raises a series of practical challenges which may limit how effective and systematic fair ML approaches can be in practice. Attempting to translate and elucidate the differences between such egalitarian theories in the context of particular machine learning tasks will likely be tricky. In simple cases, it may be that feature vectors used to train models include personal characteristics which can intuitively be classed as either chosen or unchosen (and therefore legitimate or illegitimate grounds for differential treatment according to e.g. luck egalitarianism). But more often, a contextually appropriate approach to fairness which truly captures the essence of the relevant philosophical points may hinge on factors which are not typically present in the data available in situ. Such missing data may include the protected characteristics of affected individuals, but also information relevant to an assessment of an individual's responsibility, culpability or desert–such as their socio-economic circumstances, life experience, personal development, and the relationships between them. Attempts to draw such conclusions from training data and lists of legally protected categories alone, are unlikely to do justice to the way that questions of justice arise in idiosyncratic lives and differing social contexts."
Chinese Expert Group 2019: Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence	"AI development should promote fairness and justice, protect the rights and interests of stakeholders, and promote equality of opportunity. Through continuously raising the level of technology and improving management methods, eliminate bias and discrimination in the process of data acquisition, algorithm design, technology development, product R&D, and application."

Danish Expert Group on Data ethics 2018: Data for the Benefit of the People : Recommendations from the Danish Expert Group on Data Ethics / Data i menneskets tjeneste : Anbefalinger fra Ekspertgruppen om dataetik	"Technology must not discriminate." (s.8) / "Teknologien må ikke diskriminere." (s.8)
DANSK IT's arbejdsgruppe for dataetik 2018: Dataetik : 18 dataetiske anbefalinger fra DANSK IT	"Som udgangspunkt skal en algoritme instrueres i, hvilke mål, den skal opnå, og hvilke parametre, den skal måle og analysere. Dermed indlejrer den menneskelige programmør uvægerligt en række værdier og mulige skævheder i algoritmen, som typisk vil afspejle og understøtte interesserne hos den instans eller myndighed, der driver computersystemet. Beslutninger taget på baggrund af dataprofiler og algoritmer er ikke nødvendigvis nøjagtige og fri for partiskhed eller skøn [..] Målet må derfor være, at man skal kunne inspicere algoritmer og vurdere, om deres beslutninger er tilbøjelige til at diskriminere eller på andre måder har skadelige bivirkninger for dem, der bliver vurderet af systemet." (s.24)
Den danske regering 2019: Dansk National Strategi for Kunstig Intelligens	"Kunstig intelligens må ikke reproducere fordomme, der marginaliserer befolkningsgrupper. Der skal arbejdes aktivt for at forhindre uønsket bias og fremme designs, der undgår kategorisering, som diskriminerer på baggrund af fx etnicitet, seksualitet og køn. Demografisk og faglig diversitet bør være en rettesnor i arbejdet med kunstig intelligens." (s.28)
EDPS Ethics Advisory Group 2018: Towards a digital ethics	"Like solidarity, equality is a concept with a strong political tradition in Europe and features heavily in the Charter of Fundamental Rights (Title III) in reference to equality before the law, non-discrimination, diversity, gender equality, the rights of children, the elderly and the disabled. In the digital age, novel forms of algorithmic discrimination pose a risk to equality of opportunity and to the fundamental right to be protected against digital networks that offer a wealth of often free and accessible information." (s.18)
EU Commission for the Efficiency of Justice 2018: European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment	"Specifically prevent the development or intensification of any discrimination between individuals or groups of individuals. ■Given the ability of these processing methods to reveal existing discrimination, through grouping or classifying data relating to individuals or groups of individuals, public and private stakeholders must ensure that the methods do not reproduce or aggravate such discrimination and that they do not lead to deterministic analyses or uses. ■Particular care must be taken in both the development and deployment phases, especially when the processing is directly or indirectly based on "sensitive" data. This could include alleged racial or ethnic origin, socio-economic background, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, health-related data or data concerning sexual life or sexual orientation. When such discrimination has been identified, consideration must be given to corrective measures to limit or, if possible, neutralise these risks and as well as to awareness-raising among stakeholders.

	■However, the use of machine learning and multidisciplinary scientific analyses to combat such discrimination should be encouraged." (s.9)
FAT/ML : Principles for Accountable Algorithms and a Social Impact Statement for Algorithms	“Ensure that algorithmic decisions do not create discriminatory or unjust impacts when comparing across different demographics (e.g. race, sex, etc).”
Google 2018: Objectives for AI Applications	"AI algorithms and datasets can reflect, reinforce, or reduce unfair biases. We recognize that distinguishing fair from unfair biases is not always simple, and differs across cultures and societies. We will seek to avoid unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief."
Government of the Netherlands 2019: Strategic Action Plan for Artificial Intelligence	"The prohibition of discrimination can be violated by bias in the underlying data, bias in the algorithm or by errors in classification. This could lead, for example, to unjustified distinctions being made between men and women." (s.41)
Government of UK 2019: Understanding artificial intelligence ethics and safety	"If your AI system processes social or demographic data, you should design it to meet a minimum level of discriminatory non-harm. To do this you should: - use only fair and equitable datasets (data fairness) - include reasonable features, processes, and analytical structures in your model architecture (design fairness) - prevent the system from having any discriminatory impact (outcome fairness) - implement the system in an unbiased way (implementation fairness)."
Hajian & Domingo-Ferrer / IEEE 2013: A Methodology for Direct and Indirect Discrimination Prevention in Data Mining	"At first sight, automating decisions may give a sense of fairness: classification rules do not guide themselves by personal preferences. However, at a closer look, one realizes that classification rules are actually learned by the system (e.g., loan granting) from the training data. If the training data are inherently biased for or against a particular community (e.g., foreigners), the learned model may show a discriminatory prejudiced behavior. In other words, the system may infer that just being foreign is a legitimate reason for loan denial. Discovering such potential biases and eliminating them from the training data without harming their decision-making utility is therefore highly desirable. One must prevent data mining from becoming itself a source of discrimination, due to data mining tasks generating discriminatory models from biased data sets as part of the automated decision making." (s.1445)
IEEE : A Call to Action for Businesses Using AI : Ethically Aligned Design for Business	“Make sure you understand your audience and can demonstrate value in that context. Underrepresented users (e.g., people of color, women, etc.) can be disproportionately affected when proactive prioritization of equitable AI is not built into the core strategy. Ensure that you create actionable goals that help avoid unintended bias and discrimination, and bring in subject matter experts from the affected communities as necessary.”
Leslie/The Alan Turing Institute 2019: Understanding	“All AI systems that process social or demographic data pertaining to features of human subjects must be designed to meet a minimum threshold of

artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	discriminatory non-harm. This entails that the datasets they use be equitable; that their model architectures only include reasonable features, processes, and analytical structures; that they do not have inequitable impact; and that they are implemented in an unbiased way." (s.12)
Malta AI Taskforce 2019: Towards Trustworthy AI - Malta Ethical AI Framework for public consultation	"The development, deployment, use and operation of AI systems must be fair."
Medium / Towards data science 2018: 5 Principles for Big Data Ethics	"Big data should not institutionalize unfair biases like racism or sexism. Machine learning algorithms can absorb unconscious biases in a population and amplify them via training samples."
Microsoft : Microsoft AI Principles	"AI systems should treat all people fairly."
Montreal Declaration Responsible AI : THE DECLARATIONn	"AI must be designed and trained so as not to create, reinforce or reproduce discrimination based on - among other things - social, sexual, ethnic, cultural, or religious differences."
SAP AI Ethics Steering Committee 2018: SAP's Guiding Principles for Artificial Intelligence	"Business beyond bias [...] Bias can negatively impact AI software and, in turn, individuals and our customers. This is particularly the case when there is a risk of causing discrimination or of unjustly impacting underrepresented groups. We, therefore, require our technical teams to gain a deep understanding of the business problems they are trying to solve and the data quality this demands. We seek to increase the diversity and interdisciplinarity of our teams, and we are investigating new technical methods for mitigating biases. We are also deeply committed to supporting our customers in building even more diverse businesses by leveraging AI to build products that help move business beyons bias."
SAP AI Ethics Steering Committee 2018: SAP's Guiding Principles for Artificial Intelligence	"Business beyond bias [...] Bias can negatively impact AI software and, in turn, individuals and our customers. This is particularly the case when there is a risk of causing discrimination or of unjustly impacting underrepresented groups. We, therefore, require our technical teams to gain a deep understanding of the business problems they are trying to solve and the data quality this demands. We seek to increase the diversity and interdisciplinarity of our teams, and we are investigating new technical methods for mitigating biases. We are also deeply committed to supporting our customers in building even more diverse businesses by leveraging AI to build products that help move business beyons bias."
Singapore Personal Data Protection Commission 2020: MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK second edition	"Organisations using AI in decision-making should ensure that the decision-making process is explainable, transparent and fair. Although perfect explainability, transparency and fairness are impossible to attain, organisations should strive to ensure that their use or application of AI is undertaken in a manner that reflects the objectives of these principles as far as possible. This helps build trust and confidence in AI."

Smart Dubai : ARTIFICIAL INTELLIGENCE PRINCIPLES & ETHICS	“AI systems should be fair, transparent, accountable and understandable.”
THE CENTRE FOR HUMANITARIAN DATA 2020: GUIDANCE NOTE SERIES DATA RESPONSIBILITY IN HUMANITARIAN ACTION : NOTE #4: HUMANITARIAN DATA ETHICS	"Is there a systematic skewing of the data collected and/or is there any prejudice or favoritism in the data or model? (e.g. has there been an over- or underestimation of what is being measured or are some members of the population more or less represented than others?)." (s.2)
The Japanese Society for Artificial Intelligence 2017: The Japanese Society for Artificial Intelligence Ethical Guidelines	“Members of the JSAI will always be fair. Members of the JSAI will acknowledge that the use of AI may bring about additional inequality and discrimination in society which did not exist before, and will not be biased when developing AI. Members of the JSAI will, to the best of their ability, ensure that AI is developed as a resource that can be used by humanity in a fair and equal manner.”
The Leadership Conference on Civil & Human Rights 2014: Civil Rights Principles for the Era of Big Data	“New surveillance tools and data gathering techniques that can assemble detailed information about any person or group create a heightened risk of profiling and discrimination. Clear limitations and robust audit mechanisms are necessary to make sure that if these tools are used it is in a responsible and equitable way.” “Computerized decisionmaking in areas such as employment, health, education, and lending must be judged by its impact on real people, must operate fairly for all communities, and in particular must protect the interests of those that are disadvantaged or that have historically been the subject of discrimination. Systems that are blind to the preexisting disparities faced by such communities can easily reach decisions that reinforce existing inequities. Independent review and other remedies may be necessary to assure that a system works fairly.”
The Linux Foundation : Data Values and Principles	"Recognize and mitigate bias in ourselves and in the data we use."
The Public Voice 2018: Universal Guidelines for Artificial Intelligence	“Institutions must ensure that AI systems do not reflect unfair bias or make impermissible discriminatory decisions.”
Veale & Binns 2017: Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data	"Decisions based on algorithmic, machine learning models can be unfair, reproducing biases in historical data used to train them. While computational techniques are emerging to address aspects of these concerns through communities such as discrimination-aware data mining (DADM) and fairness, accountability and transparency machine learning (FATML), their practical implementation faces real-world challenges. For legal, institutional or commercial reasons, organisations might not hold the data on sensitive attributes such as gender, ethnicity, sexuality or disability needed to diagnose and mitigate emergent indirect discrimination-by-proxy, such as redlining. Such organisations might also lack the knowledge and capacity to identify and

	manage fairness issues that are emergent properties of complex sociotechnical systems. This paper presents and discusses three potential approaches to deal with such knowledge and information deficits in the context of fairer machine learning. Trusted third parties could selectively store data necessary for performing discrimination discovery and incorporating fairness constraints into model-building in a privacy-preserving manner. Collaborative online platforms would allow diverse organisations to record, share and access contextual and experiential knowledge to promote fairness in machine learning systems. Finally, unsupervised learning and pedagogically interpretable algorithms might allow fairness hypotheses to be built for further selective testing and exploration. Real-world fairness challenges in machine learning are not abstract, constrained optimisation problems, but are institutionally and contextually grounded. Computational fairness tools are useful, but must be researched and developed in and with the messy contexts that will shape their deployment, rather than just for imagined situations. Not doing so risks real, near-term algorithmic harm." (s.1)
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	"In the absence of relevant differences between two or more situations, consistency requires that the same standards be applied across them. While consistency in decision-making may be regarded as valuable in its own right, adherence to a practice of consistency may help actors to secure other values, such as fairness and trustworthiness." (s.246)
Gennemsigtighed	
Accenture 2016: Universal principles of data ethics : 12 guidelines for developing ethics codes	"Maximizing transparency at the point of data collection can minimize more significant risks as data travels through the data supply chain." (s.9)
ACM US Public Policy Council 2017: Statement on Algorithmic Transparency and Accountability	"Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions." "A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals." "Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts."
Amsterdam Economic Board (TADA): The 6 principles of our manifesto	"What types of data are collected? For what purpose? And what are the outcomes and results? We are always transparent about those aspects."

Australian Government : AI Ethics Principles	"There should be transparency and responsible disclosure to ensure people know when they are being significantly impacted by an AI system, and can find out when an AI system is engaging with them." "This principle aims to ensure responsible disclosure when an AI system is significantly impacting on a person's life. The definition of the threshold for 'significant impact' will depend on the context, impact and application of the AI system in question."
Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	"Related to the issue of control over public identity formation, is the valid concern that children may not have full knowledge or understanding of the implications of data accessibility and subsequent uses."
Bertelsmann Stiftung & iRights.Lab 2019: Algo.Rules : Design rules for algorithmic systems	"The decision-making processes within an algorithmic system must always be comprehensible." (s.5) "In order to question and review decisions resulting from an algorithmic system, people must be able to understand both direct and indirect effects of the system as well as how decisions are reached. Information about the data and models on which the system is based, its architecture and potential effects must be published in easily understood terms. In addition, it is important to check whether an objective can be achieved without a significant loss in quality through the use of a less complex algorithmic system that involves an easier to understand mode of operation." (ibid.) "The use of an algorithmic system must be identified as such." (s.5) "People interacting with algorithmic systems must be able to identify that a decision or prediction is based on an algorithm. This is particularly important in cases where the system imitates a human being in how it interacts (e.g., through language or appearance)." (ibid.)
DANSK IT's arbejdsgruppe for dataetik 2018: Dataetik : 18 dataetiske anbefalinger fra DANSK IT	"Cathy O'Neil påpeger, at algoritmer er tilbøjelige til at forstærke de eksisterende magtstrukturer, fordi algoritmerne kan bruges til at automatisere behandlingen af den svageste og mindst velstående del af befolkningen. Computersystemerne er sorte bokse, og det er kun et fåtal, der overhovedet har den tekniske indsigt og viden til at forstå, hvordan algoritmerne når frem til deres konklusioner, når de f.eks. laver kreditvurderinger eller frasorterer jobansøgninger [...] For nylig efterlyste EU-Kommissionen i en rapport om etik og kunstig intelligens, at algoritmer skal være explainable, dvs. at man skal kunne inspicere en algoritme og gennemskue, hvordan den når frem til sine konklusioner. Rapporten argumenterede for, at vi alle har en ret til forklaring af formål og anvendelse af en teknologi, fremfor bare en ret til underretning, hvis der sker noget. Det handler om at gøre det muligt at forklare og argumentere for formålet med en algoritme, da algoritmer langt fra er neutrale og nøgterne, som man skulle tro." (s.24)
Den danske regering 2019: Dansk National Strategi for Kunstig Intelligens	"Forklarlighed indebærer, at man kan beskrive, kontrollere og genskabe data, bagvedliggende logikker og konsekvenser af anvendelsen af kunstig intelligens, fx ved at kunne spore og forklare beslutninger og beslutningsunderstøttelse. Forklarlighed er ikke ensbetydende med fuld transparens omkring algoritmer, da der blandt andet er forretningsmæssige hensyn i den private sektor. Offentlige myndigheder har dog et særligt ansvar for at sikre åbenhed og gennemsigtighed ved brug af algoritmer" (s.28)

Drew 2016: Data science ethics in government	"Be as open and accountable as possible without putting people at risk [...] Transparency is essential to make the case for the benefits of data science and to avoid accusation of nefarious 'secret' big data projects. It is also a good antiseptic for unethical behaviour. Ideally, people would like transparency at all stages of a data science project, being told when and why data are being collected about them as well as whether the outcome of the data science project is achieved." (s.7)
EU Commision for the Efficiency of Justice 2018: European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment	"Make data processing methods accessible and understandable, authorise external audits: ■A balance must be struck between the intellectual property of certain processing methods and the need for transparency (access to the design process), impartiality (absence of bias), fairness and intellectual integrity (prioritising the interests of justice) when tools are used that may have legal consequences or may significantly affect people's lives. It should be made clear that these measures apply to the whole design and operating chain as the selection process and the quality and organisation of data directly influence the learning phase. ■The first option is complete technical transparency (for example, open source code and documentation), which is sometimes restricted by the protection of trade secrets. The system could also be explained in clear and familiar language (to describe how results are produced) by communicating, for example, the nature of the services offered, the tools that have been developed, performance and the risks of error. Independent authorities or experts could be tasked with certifying and auditing processing methods or providing advice beforehand. Public authorities could grant certification, to be regularly reviewed." (s.11)
FAT/ML : Principles for Accountable Algorithms and a Social Impact Statement for Algorithms	"Ensure that algorithmic decisions as well as any data driving those decisions can be explained to end-users and other stakeholders in non-technical terms." "Enable interested third parties to probe, understand, and review the behaviour of the algorithm through disclosure of information that enables monitoring, checking, or criticism, including through provision of detailed documentation, technically suitable APIs, and permissive terms of use."
Forsikring & Pension : Cool eller creepy? : Databrug og dataetiske principper i forsikrings- og pensionsbranchen	"Transparens er et nøgleprincip for forsikrings- og pensionsbranchen. Transparens om databrug giver dig mulighed for selv at kontrollere, hvad dine data bliver brugt til, og om de data, der bliver lagt til grund, er rigtige [...] Det skal være klart, hvad du siger ja til, hvorfor og hvordan dine data har betydning for prisen, bliver opbevaret, og hvad de bliver anvendt til. Selskaberne vil samtidig fortælle åbent, hvordan de arbejder med dataetik. Det kan fx ske i deres årsberetning, som en del af deres CSR-politik eller på deres hjemmeside." (s.8)
Forsikring & Pension 2019: Databrug og dataetik : Dilemmaer og mulige positioner for forsikrings- og pensionsbranchen	"Ved jeg, hvor mine data ligger, hvad de bliver brugt til og hvordan? Ved jeg, hvad jeg siger ja og nej til, når data indsamles om mig? Og kender jeg i et rimeligt omfang konsekvenserne af at træffe beslutning om en given datadeling?"

Future of Life 2017: ASILOMAR AI PRINCIPLES	“Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.” "Failure transparency [...] If an AI system causes harm, it should be possible to ascertain why."
Government of Canada: Our guiding principles	"Be as open as we can be by sharing source code, training data, and other relevant information, all while protecting personal information, system integration, and national security and defence." "Provide meaningful explanations about AI decision making, while also offering opportunities to review results and challenge these decisions." "Be transparent about how and when we are using AI, starting with a clear user need and public benefit."
Government of the Netherlands 2019: Strategic Action Plan for Artificial Intelligence	"This right applies, among other things, to automated decision-making. If it is not transparent whether, and if so which, algorithms have been used to make or prepare a decision, or what assumptions and data form the basis of this decision, there will be pressure on the contestability and substantiation of decisions and statements, and on the balance between the parties to the proceedings (equality of arms)." (s.41)
Government of UK 2019: Understanding artificial intelligence ethics and safety	"Designers and implementers of AI systems should be able to: 1) Explain to affected stakeholders how and why a model performed the way it did in a specific context. 2) Justify the ethical permissibility, the discriminatory non-harm and the public trustworthiness of its outcome and of the processes behind its design and use."
IBM 2019: Everyday Ethics for Artificial Intelligence	"AI should be designed for humans to easily perceive, detect, and understand its decision process." (s.26) "In general, we don’t blindly trust those who can’t explain their reasoning. The same goes for AI, perhaps even more so. As an AI increases in capabilities and achieves a greater range of impact, its decision-making process should be explainable in terms people can understand. Explainability is key for users interacting with AI to understand the AI’s conclusions and recommendations. Your users should always be aware that they are interacting with an AI. Good design does not sacrifice transparency in creating a seamless experience. Imperceptible AI is not ethical AI." (s.26) Til prinсippet hører også "Recommended actions to take" og eksempler (se s. 28-30).
ITI (Information Technology Industry Council) : AI Policy Principles : Executive Summary	"We are committed to partnering with others across government, private industry, academia, and civil society to find ways to mitigate bias, inequity, and other potential harms in automated decision-making systems. Our approach to finding such solutions should be tailored to the unique risks presented by the specific context in which a particular system operates. In many contexts, we believe tools to enable greater interpretability will play an important role." (s.3)
Leslie/The Alan Turing Institute 2019: Understanding	“Designers and implementers of AI systems must be able (1) to explain to affected stakeholders in everyday language how and why a model performed

artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	the way it did in a specific context and (2) to justify the ethical permissibility, the discriminatory non-harm, and the public trustworthiness both of its outcome and of the processes behind its design and use." (s.12)
Malta AI Taskforce 2019: Towards Trustworthy AI - Malta Ethical AI Framework for public consultation	"End-users and other members of the public should be able to understand and challenge the operation of AI systems, as required for the particular use case."
Medium / Towards data science 2018: 5 Principles for Big Data Ethics	"Customers should have a transparent view of how our data is being used or sold, and the ability to manage the flow of their private information across massive, third-party analytical systems."
Microsoft : Microsoft AI Principles	"AI systems should be understandable."
Mittelstadt & Floridi 2016: The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts	"Big Data is increasingly becoming the sole domain of large organisations, despite calls to allow data subjects to benefit from and manipulate their data. This situation can be troublesome for several reasons, foremost due to the inability of 'underprivileged' individual data subjects and organisations both to understand and have access to the methods, logic or at least ''decisional criteria'' behind Big Data analysis and decision-making processes. Furthermore, it is often unclear which individuals and organisations can access or buy one's data."
Montreal Declaration Responsible AI: THE DECLARATION	"AI processes that make decisions affecting a person's life, quality of life or reputation must be intelligible to their creators. The decisions made by AI affecting a person's life, quality of life, or reputation should always be justifiable in a language that is understood by the people who use them or who are subjected to the consequences of their use. Justification consists in making transparent the most important factors and parameters shaping the decision, and should take the same form as the justification we would demand of a human making the same kind of decision. The code for algorithms, whether public or private, must always be accessible to the relevant public authorities and stakeholders for verification and control purposes. The discovery of AI operating errors, unexpected or undesirable effects, security breaches, and data leaks must imperatively be reported to the relevant public authorities, stakeholders, and those affected by the situation. In accordance with the transparency requirement for public decisions, the code for decision-making algorithms used by public authorities must be accessible to all, with the exception of algorithms that present a high risk of serious danger if misused. For public AI that have a significant impact on the life of citizens, citizens should have the opportunity and skills to deliberate on the social parameters of these AI, their objectives, and the limits of their use. We must at all times be able to verify that AI are doing what they were programmed for and what they are used for. Any person using a service should know if a decision concerning them or affecting them was made by an AI. AI research should remain open and accessible to all."

Norwegian Ministry of Local Government and Modernisation 2020: National Strategy for Artificial Intelligence	"Decisions made by systems built on AI must be traceable, explainable and transparent. This means that individuals or legal persons must have an opportunity to gain insight into how a decision that affects them was made. Traceability facilitates auditability as well as explainability. Transparency is also about computer systems not pretending to be human being; human beings must have the right to know if they are interacting with an AI system."
OECD 2019: Recommendation of the Council on Artificial Intelligence	"AI actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art: 1) to foster a general understanding of AI systems 2) To make stakeholders aware of their interactions with Ai systems, including in the workplace 3) To enable those affected by an AI system to understand the outcome 4) To enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision."
Regulatory and Ethics Working Group of the Global Alliance for Genomics and Health 2014: Framework for Responsible Sharing of Genomic and Health-Related Data	"• Develop clearly defined and accessible information on the purposes, processes, procedures and governance frameworks for data sharing. Such information should be presented in a way that is understandable and accessible in both digital and non-digital formats. • Provide clear information on the purpose, collection, use and exchange of genomic and health-related data, including, but not limited to: data transfer to third parties; international transfer of data; terms of access; duration of data storage; identifiability of individuals and data and limits to anonymity or confidentiality of data; communication of results to individuals and/or groups; oversight of downstream uses of data; commercial involvement; proprietary claims; and processes of withdrawal from data sharing. • Implement procedures for fairly determining requests for data access and/or exchange." (s.4)
Richards & King 2014: Big Data Ethics	"Transparency, like confidentiality, also fosters trust by being able to hold others accountable [...]" (s.419) "Transparency has heightened importance with the arrival of big data. The power of big data comes in large part from secondary uses of data sets to produce new predictions and inferences." (s.421) "Transparency inherently includes a tension between openness and secrecy. This tension can cause paradoxes. Transparency of sensitive corporate or government secrets could harm important interests, such as trade secrets or national security. Too little transparency can lead to unexpected outcomes and a lack of trust. Transparency also carries the risk that inadvertent disclosures will cause unexpected outcomes that harm privacy and breach confidentiality." (s.420)
SAP AI Ethics Steering Committee 2018: SAP's Guiding Principles for Artificial Intelligence	"Our systems are held to specific standards in accordance with their level of technical ability and intended usage. Their input capabilities, intended purpose, and limitations will be communicated clearly to our customers, and we provide means for oversight and control by customers and users. They are, and will always remain, in control of the deployment of our products. We

	actively support industry collaboration and will conduct research to further system transparency.”
Slade & Prinsloo 2013: Learning analytics: ethical issues and dilemmas	"Important for learning analytics as moral practice is that higher education institutions should be transparent regarding the purposes for which data will be used, under which conditions, who will have access to data and the measures through which individuals’ identity will be protected. The assumption that participating in public online forums provides blanket permission for use of data should not be acceptable. Higher education institutions have an obligation to protect participant data on the institutional LMS, and also to inform students of possible risks when teaching and learning occurs outside the boundaries of institutional jurisdiction." (s.14)
THE CENTRE FOR HUMANITARIAN DATA 2020: GUIDANCE NOTE SERIES DATA RESPONSIBILITY IN HUMANITARIAN ACTION : NOTE #4: HUMANITARIAN DATA ETHICS	"Is there clear documentation of the data management process and visibility on how the model or algorithm(s) function? (e.g. can someone not directly involved in the process explain what is happening?)" (s.2) "Are the rights to the data and related insights derived clearly defined? (e.g. is it clear how decisions are made regarding how and by whom the data can be used, how problems in or related to the data are rectified, and other related issues?)" (s.2)
The Japanese Society for Artificial Intelligence 2017: The Japanese Society for Artificial Intelligence Ethical Guidelines	“Members of the JSAI must aim to improve and enhance society's understanding of AI. Members of the JSAI understand that there are diverse views of AI within society, and will earnestly learn from them. They will strengthen their understanding of society and maintain consistent and effective communication with them, with the aim of contributing to the overall peace and happiness of mankind.”
The Linux Foundation : Data Values and Principles	"Respect and invite fair criticism while promoting the identification and open discussion of errors, risks, and unintended consequences of our work."
The Public Voice 2018: Universal Guidelines for Artificial Intelligence	“No institution shall establish or maintain a secret profiling system.” “All individuals have the right to know the basis of an AI decision that concerns them. This includes access to the factors, the logic, and techniques that produced the outcome.”
The Task Force on Artificial Intelligence of the Agency for Digital Italy 2018: White Paper on Artificial Intelligence at the service of citizens	"The issue of the responsibility of public administration also has to do with the duties of the latter with respect to citizens, when it decides to provide them with services or to make decisions that concern them, using Artificial Intelligence solutions. The functioning of the latter must meet criteria of transparency and openness. Transparency becomes a fundamental prerequisite to avoid discrimination and solve the problem of information asymmetry, guaranteeing citizens the right to understand public decisions. It is also necessary to think about the policies chosen to determine the benchmarks (benchmark policy) to avoid effects of a larger scale: as an administrator can act in a non-transparent manner by pursuing not the common good but private interests, a non-transparent algorithm could realize the same offenses even more broadly, producing not only injustices but also social discrimination." (s.37)

Tranberg, Hasselbalch, Olsen & Byrne / DataEthics.eu - The independent Thinkdotank 2018: DATAETHICS – Principles and Guidelines for Companies, Authorities & Organisation	"Data processing activities and automated decisions must make sense for the individual. They must be truly transparent and explainable. The purpose and interests of data processing must be clearly understood by the individual in terms of understanding risks, as well as social, ethical and societal consequences." (s.10)
UK Government, Department for Digital, Culture, Media & Sport 2018: Guidance : Data Ethics Framework	"Make your work transparent and be accountable [...] You should be transparent about the tools, data and algorithms you used to conduct your work, working in the open where possible. This allows other researchers to scrutinise your findings and citizens to understand the new types of work we are doing." (s.5)
UK Statistics Authority : Data Ethics	"The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public."
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	"Governments should provide regular reports about their use of AI in policing, intelligence, and security." "AI should be explainable, able to provide insight into its functioning." "The data used to train AI systems should be transparent."
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	"Transparency is openness to public scrutiny of decision-making, processes, and actions. Transparency helps to demonstrate respect for persons and contributes to trustworthiness." (s.246) "Trustworthiness is the property of being worthy of trust. It is a value that applies not only to individuals, organisations, governments, and institutions, but also to data, evidence, and systems. It can manifest procedurally as being transparent and truthful, reliable and consistent, or dependable." (s.246)
Godgørenhed	
Accenture 2016: Universal principles of data ethics : 12 guidelines for developing ethics codes	"When insights derived from data could impact the human condition, the potential harm to individuals and communities should be the paramount consideration. Big data can produce compelling insights about populations, but those same insights can be used to unfairly limit an individual's possibilities." (s.8)
Australian Government : AI Ethics Principles	"Throughout their lifecycle, AI systems should benefit individuals, society and the environment." "This principle aims to clearly indicate from the outset that AI systems should be used for beneficial outcomes for individuals, society and the environment. AI system objectives should be clearly identified and justified. AI systems that help address areas of global concern should be encouraged, like the United Nation's Sustainable Development Goals. Ideally, AI systems should be used to benefit all human beings, including future generations. AI systems designed for legitimate internal business purposes, like increasing efficiency, can have broader impacts on individual, social and environmental wellbeing. Those impacts, both positive and negative, should be accounted for throughout the AI system's lifecycle, including impacts outside the organisation."

Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	"A further concern applicable to adults but with particular salience for children, is the potential for decision makers to substitute direct dialogue and engagement with children with the cheaper and quicker approach of passive big data collection." (s.17)
Brakewood & Poldrack 2013: The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data	“Beneficence, in the Belmont report, is a positive obligation to both not injure a participant (not limited only to physical injury), but also to maximize benefits and minimize necessary harms (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978) […] It is the responsibility of an ethical investigator to ensure that subjects receive the maximum benefit to participation while those harms are minimized.” (s.673)
Chinese Expert Group 2019: Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence	"AI should: promote green development and meet the requirements of environmental friendliness and resource conservation; promote coordinated development, push forward the transformation and upgrading of all walks of life, and narrow regional disparities; promote inclusive development, strengthen AI education and popularization of science, improve the adaptability of disadvantaged groups, and strive to erase the digital divide; promote shared development, avoid data and platform monopolies, and encourage open and orderly competition." "Respect the natural laws of AI development; while promoting the innovative and orderly development of AI, search for and resolve risks that might arise. Continuously upgrade intelligent technological methods, optimize management mechanisms, perfect governance systems, and promote governance principles throughout the entire life cycle of AI products and services. Continue to research and anticipate potential future risks from increasingly advanced AI, and ensure that AI always moves in a direction that is beneficial to society."
Drew 2016: Data science ethics in government	"Data science projects should always start with a clear policy or operational need, which is an important first point for two reasons. Firstly, participants were more willing to support data science projects when there is a clear public benefit (and when they can see the value of data science over more traditional methods). And secondly, the value of the public benefit affects how much risk participants were willing to take in the design of the data science, for example in the type of data used (principle 2) or the type or volume of intended or unintended consequences that might occur within the model (principle 3)."
Fairfield & Shtein 2014: Big Data, Big Problems : Emerging Issues in the Ethics of Data Science and Journalism	"Beneficence requires that the researcher minimize harm where possible and relate it proportionally to the potential benefit of the study. As the Belmont Report notes, “beneficence is understood as an obligation.” Two complementary rules have been recognized as falling under the beneficence category: “(1) do not harm and (2) maximize possible benefits and minimize possible harms.”" (s.40)
Future of Life 2017: ASILOMAR AI PRINCIPLES	“AI technologies should benefit and empower as many people as possible.” “The economic prosperity created by AI should be shared broadly, to benefit all of humanity.”

Google 2018: Objectives for AI Applications	"The expanded reach of new technologies increasingly touches society as a whole. Advances in AI will have transformative impacts in a wide range of fields, including healthcare, security, energy, transportation, manufacturing, and entertainment. As we consider potential development and uses of AI technologies, we will take into account a broad range of social and economic factors, and will proceed where we believe that the overall likely benefits substantially exceed the foreseeable risks and downsides. AI also enhances our ability to understand the meaning of content at scale. We will strive to make high-quality and accurate information readily available using AI, while continuing to respect cultural, social, and legal norms in the countries where we operate. And we will continue to thoughtfully evaluate when to make our technologies available on a non-commercial basis."
Herschel & Miori 2017: Ethics & Big Data	"Unlike Kantianism, assessing the ethics of Big Data from a Utilitarian perspective is fraught with complications. It requires that acts and rules be assessed using a utilitarian calculus where the god and bad of Big Data are weighed on a scale. From an Act Utilitarian perspective, for example, one would have to quantify the plusses and minuses of Big Data consequences relative to such factors as the intensity of the experience, its duration, the probability that something would occur, how close the experiences are in space and time, its ability to produce more experiences of the same kind, the extent to which pleasure is not diluted by pain or vice versa, and the number of people affected. To make a decision as to whether a use of Big Data is right or wrong, one would total the positive and negative consequences to all being affected, total up the positives and the negatives and choose the alternative with the highest amount. Rule Utilitarianism is more simplistic than Act Utilitarianism. It argues that we should follow a moral rule because its adoption would result in the greatest net increase in happiness. Big Data would be assessed relative to the weighing of its harms and benefits to society."
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	"• Prioritise social welfare, public interest, and the consideration of the social and ethical impacts of innovation in determining the legitimacy and desirability of AI technologies • Use AI to empower and to advance the interests and well-being of as many individuals as possible • Think big-picture about the wider impacts of the AI technologies you are conceiving and developing. Think about the ramifications of their effects and externalities for others around the globe, for future generations, and for the biosphere as a whole." (s.11)
Malta AI Taskforce 2019: Towards Trustworthy AI - Malta Ethical AI Framework for public consultation	"AI systems should not cause harm at any stage of their lifecycle to humans, the natural environment or other living beings."
Norwegian Ministry of Local Government and Modernisation 2020: National Strategy for Artificial Intelligence	"AI must be developed with consideration for society and the environment, and must have no adverse effects on institutions, democracy or society at large."

OECD 2019: Recommendation of the Council on Artificial Intelligence	“Stakeholders should proactively engage in responsible stewardship of trustworthy Ai in pursuit in beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender, and other inequalities and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.”
Regulatory and Ethics Working Group of the Global Alliance for Genomics and Health 2014: Framework for Responsible Sharing of Genomic and Health-Related Data	"• Consider the realistic harms and benefits of data sharing on and with individuals, families and communities, including opportunity costs associated with both sharing and not sharing data. Potential realistic benefits may include development of new scientific knowledge and applications, enhanced efficiency, reproducibility and safety of research projects or processes, and more informed decisions about health care. Potential realistic harms may include invasions of privacy or breach of confidentiality and invalid conclusions about research projects. • Conduct data sharing with a view towards minimizing harms and maximizing benefits to not just those who contribute their data, but also to society and health care systems as a whole, particularly where data pertains to disadvantaged people. Benefits arising from data sharing may not be uniformly distributed throughout communities around the world and may depend on the usability of data within a specified context, national priorities, as well as a specific community’s concern about health and interpretation of wellbeing. • Undertake a proportionate assessment of the benefits and risks of harm in data sharing, which is periodically monitored according to the reasonable foreseeability of such harms and benefits. Such an assessment may also incorporate mechanisms that track subsequent harms, should they materialize, so as to help inform future policy." (s.5)
Taylor 2016: The ethics of big data as a public good: which public? Whose good?	"There are two possible definitions of a public good that are relevant in this scenario [data philantrophy case]. One is the notion that data should be made available to help international organizations promote social good in the public sphere, i.e. that data should be more public to create more good impacts from it. The second is the argument that data should be formally defined as a public good, because its potential power to fight poverty and disease and to inform emergency response is such that international institutions should have access to it. This is effectively an argument for reordering power relations with regard to digital data. The public good argument with regard to information and knowledge has been defined by information economists Stiglitz and Varian. In their definition, due to the low costs of reproduction, knowledge is a resource that is both non-rivalrous (there is no extra cost incurred when others use it) and non-excludable (it is impossible to keep others from using it). However, in this definition, there are caveats with relevance to big data: Stiglitz and Varian warn that knowledge can be made functionally excludable where the private sector gains value from controlling it, and that regimes also determine the extent to which it is excludable, for example, in the form of taxes and patents. Purtova brings this debate up to date, identifying digital data deriving from people as a ‘system resource’ comprising an ecosystem of people, platforms and profiles, and concluding that while it may be possible for knowledge to be a public good, it is not possible to make the same claim for digital data. In fact, the language of the ‘personal data ecosystem’ is already in use by the World

	Economic Forum, among others, to explain the ways in which the knowledge produced through digital data is inherently commercial, and operates as an interaction between individuals and firms." (s.1)
The Linux Foundation : Data Values and Principles	"Use data to improve life for our users, customers, organizations, and communities."
UK Government, Department for Digital, Culture, Media & Sport 2018: Guidance : Data Ethics Framework	"Using data in more innovative ways has the potential to transform how public services are delivered. We must always be clear about what we are trying to achieve for users - both citizens and public servants." (s.5)
UK Statistics Authority : Data Ethics	"The use of data has clear benefits for users and serves the public good." "The views of the public are considered in light of the data used and the perceived benefits of the research."
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	"For all AI applications, the potential benefits need to be balanced against the environmental impact of the entire AI and IT production cycle."
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	"Harm minimisation involves reducing the possibility of real or perceived harms (physical, economic, psychological, emotional, or reputational) to persons." (s.245) "Public benefit is the overall good that society as a whole receives from a given project. This includes consideration of effects on wellbeing, distribution, societal cohesion, human rights, and other sources of value to society. It may not be possible to measure these factors by the same standards, so some judgement and critical analysis will be required in determining what is publicly beneficial." (s.245) "Stewardship reflects a relationship with things, such as data, to promote twin objectives of taking care of the object of attention as well as seeking actively to promote its value and utility. It involves guiding others with prudence and care across one or more endeavours—without which there is risk of impairment or harm—and with a view to collective betterment." (s.245)
Zimmer 2018: Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity	"Since subjects may be exposed to risks or experience harm during, or because of, a research study, a core principle of research ethics is non-maleficence—the duty to avoid, prevent, or minimize harms to subjects. Research subjects must not be subjected to unnecessary risks of harm, and their participation in research must be essential to achieving scientifically and societally important aims that cannot be realized without the participation of human subjects. Put most simply, research should not harm participants, and ethical research practices must work toward minimizing any risk of harm. There are numerous types of harm that participants might be subjected to, including physical harm, psychological distress, social and reputational disadvantages, harm to one's financial status, and breaches of one's expected privacy, confidentiality, or anonymity. To minimize the risk of these harms, research ethics guidelines typically point to other key principles and operational practices, including obtaining informed consent and protecting the privacy and confidentiality of participants." (s.3)

Lighed	
Accenture 2016: Universal principles of data ethics : 12 guidelines for developing ethics codes	"While everyone deserves the social and economic benefits of data, not everyone is equally impacted by the processes of data collection, correlation, and prediction. Data professionals should strive to mitigate the disparate impacts of their products and listen to the concerns of affected communities." (s.9)
Amsterdam Economic Board (TADA): The 6 principles of our manifesto	"Our digital city is inclusive. We take into account the differences between individuals and groups, without losing sight of equality." "Data that government authorities, companies and other organizations generate from the city and collect about the city are held in common. Everyone can use them. Everyone can benefit from them. We make mutual agreements about this."
Danish Expert Group on Data ethics 2018: Data for the Benefit of the People : Recommendations from the Danish Expert Group on Data Ethics / Data i menneskets tjeneste : Anbefalinger fra Ekspertgruppen om dataetik	"When developing technological solutions, involve as many trade groups of different genders, ages, ethnicities, etc. as possible." (s.8) / "Flest mulige faggrupper med forskellig køn, alder, etnicitet osv. skal involveres i udviklingen af teknologiske løsninger." (s.8)
EDPS Ethics Advisory Group 2018: Towards a digital ethics	"Solidarity refers to a relation to others, the unity of community values, aims, interests, objectives or standards, past, present and future. In its most elementary form, solidarity corresponds to something shared, something that holds a group together in an environment. Solidarity has played a key role in the geopolitical discourse of European construction from its very origin. It is a core concept of Title IV of the Charter of Fundamental Rights, which contains guiding provisions about societal security, health care, access to economic services and consumer protection, to name a few." (s.18)
Forsikring & Pension : Cool eller creepy? : Databrug og dataetiske principper i forsikrings- og pensionsbranchen	"Data kan skabe mulighed for individuelt tilpassede produkter og mere præcis rådgivning. Den øgede mængde datapunkter, der findes om dig, kan bruges til mere effektivt at fastsætte risikoen ved at forsikre dig og dine ting og give dig en mere præcis pris [...] Men det er et grundprincip i forsikrings og pensionsbranchen, at alle skal have adgang til forsikringer på rimelige vilkår, og at der også er løsninger til de særligt sårbare, der kan blive ramt af eksklusion grundet stigende personalisering og individuelle priser [...] vigtigt, at det enkelte selskab træffer etiske valg og sikrer, at adfærdspåvirkning ved brug af data sker i kundernes og medlemmernes interesse, fx til at forebygge skader eller gøre det nemmere at træffe de rigtige valg." (s.10)
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and	"• Safeguard the integrity of interpersonal dialogue, meaningful human connection, and social cohesion • Prioritise diversity, participation, and inclusion at all points in the design, development, and deployment processes of AI innovation.

implementation of AI systems in the public sector	• Encourage all voices to be heard and all opinions to be weighed seriously and sincerely throughout the production and use lifecycle • Use the advancement and proliferation of AI technologies to strengthen the developmentally essential relationship between interacting human beings. • Utilise AI innovations pro-socially so as to enable bonds of interpersonal solidarity to form and individuals to be socialised and recognised by each other • Use AI technologies to foster this capacity to connect so as to reinforce the edifice of trust, empathy, reciprocal responsibility, and mutual understanding upon which all ethically well-founded social orders rest." (s.10) "• Treat all individuals equally and protect social equity • Use digital technologies as an essential support for the protection of fair and equal treatment under the law." (s.11)
Montreal Declaration Responsible AI : THE DECLARATION	"AI should help improve risk management and foster conditions for society with a more equitable and mutual distribution of individual and collective risks." "AI development must help eliminate relationships of domination between groups and people based on differences of power, wealth, or knowledge. AI development must produce social and economic benefits for all by reducing social inequalities and vulnerabilities. Industrial AI development must be compatible with acceptable working conditions at every step of their life cycle, from natural resources extraction to recycling, and including data processing. The digital activity of users of AI and digital services should be recognized as labor that conributes to the functioning of algorithms and creates value. Access to fundamental resources, knowledge and digital tools must be guaranteed for all. We should support the development of commons algorithms - and of open data needed to train them - and expand their use, as a socially equitable objective. AI development environments, whether in research or industry, must be inclusive and reflect the diversity of the individuals and groups of the society."
Norwegian Ministry of Local Government and Modernisation 2020: National Strategy for Artificial Intelligence	"When developing and using AI, it is especially important to ensure that AI contribute to inclusion and equality, and that discrimination be avoided. Datasets that are used to train AI systems can contain historical bias, be incomplete or incorrect. Identifiable and discriminatory bias should, if possible, be removed in the collection phase. Selection bias can be counteracted by putting in place oversight processes to analyse and correct the system's decisions in light of the purpose."
The Linux Foundation : Data Values and Principles	"Build teams with diverse ideas, backgrounds, and strengths." "Maximize diversity, connectivity, and accessibility among data projects, collaborators, and outputs."
Tranberg, Hasselbalch, Olsen & Byrne / DataEthics.eu - The independent Thinkdotank 2018: DATAETHICS – Principles and Guidelines for Companies, Authorities & Organisation	"Democratic data processing is based on an awareness of the societal power relations that data systems sustain, reproduce or create. When processing data, special attention should be paid to vulnerable people, who are particularly vulnerable to profiling that may adversely affect their self-determination and control or expose them to discrimination or stigmatisation, for example due to their financial, social or health related conditions. Paying attention to vulnerable people also involves working actively to reduce bias in the development of self-learning algorithms." (s.11-12)

UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	“AI should be inclusive, aiming to avoid bias and allowing for diversity and avoiding a new digital divide.”
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	"Solidarity is the commitment among persons with recognised morally relevant sameness or similarity to sharing costs and benefits for the good of a group, community, nation, or global population." (s.245)
Privatliv	
Accenture 2016: Universal principles of data ethics : 12 guidelines for developing ethics codes	"Data subjects hold a range of expectations about the privacy and security of their data and those expectations are often context-dependent. Designers and data professionals should give due consideration to those expectations and align safeguards and expectations as much as possible." (s.8) "Privacy does not mean secrecy, as private data might need to be audited based on legal requirements, but that private data obtained from a person with their consent should not be exposed for use by other businesses or individuals with any traces to their identity." "Third party companies share sensitive data — medical, financial or locational — and need to have restrictions on whether and how that information can be shared further."
Australian Government : AI Ethics Principles	"Throughout their lifecycle, AI systems should respect and uphold privacy rights and data protection, and ensure the security of data." "This principle aims to ensure respect for privacy and data protection when using AI systems. This includes ensuring proper data governance, and management, for all data used and generated by the AI system throughout its lifecycle. For example, maintaining privacy through appropriate data anonymisation where used by AI systems."
Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	"[...] most experts agree that data anonymization is not foolproof and that there is a tension between utility and anonymity: Data can often be either useful, or anonymous, rarely both. Techniques such as ‘differential privacy’ can go some way to ensure privacy protection and prevention of misuse of data, but much greater regulation is needed in this area. This capacity for data to be re-identifiable has the potential to impact children throughout their life cycle, in negative ways." (s.16)
Chinese Expert Group 2019: Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence	"AI development should respect and protect personal privacy and fully protect the individual's right to know and right to choose. In personal information collection, storage, processing, use, and other aspects, boundaries should be set and standards should be established. Improve personal data authorization and revocation mechanisms to combat any theft, tampering, disclosure, or other illegal collection or use of personal information."
DANSK IT’s arbejdsgruppe for dataetik 2018: Dataetik : 18 dataetiske anbefalinger fra DANSK IT	"Privacy by design [...] privacy by default [...]. Formålet med begge principper er at minimere mængden af data, der automatisk deles, og sørge for, at forbrugerne som udgangspunkt selv har ret til at bestemme, hvor meget af deres data, der skal deles mv. [...] Databeskyttelse gennem design skabes på basis af principper som **proportionalite**t, hvor mål og metode skal være i

	balance, **beskyttelse og kontrol** ved at have gennemsigtighed og ansvarlighed i forbindelse med processer og **inddragelse og involvering** ved at være åben og transparent i forhold til produkt og flow." (s.20)
Datenethik Kommission 2019: Opinion of the Data Ethics Commission	"The protection of human dignity and self-determination are closely and materially linked with the protection of privacy. The individual's right to determine who may access which personal information relating to him or her, and when and for what purpose they may do so, is justified by the supreme ethical importance of the ability to prevent intrusions into one's private sphere and also to appear in public in the certainty that one's privacy is protected. Efforts to protect human dignity must include legislative measures to regulate the responsible use of personal data. A further aspect of privacy is the need to preserve the integrity of an individual's personal identity. For example, this integrity may be violated if an algorithmic system – using data collected for entirely different purposes – "calculates" the personality of an individual together with his or her preferences and proclivities, and the system operator then uses these calculations for its own purposes, regardless of or even contrary to the individual's will." (s.45)
Drew 2016: Data science ethics in government	"The government should always use the minimum data necessary to achieve the public benefit." (s.5)
Forsikring & Pension 2019: Databrug og dataetik : Dilemmaer og mulige positioner for forsikrings- og pensionsbranchen	"Har jeg kontrol over mine data, og hvem kigger med i mit privatliv, og til hvilke formål de bliver brugt?" "Kan jeg være sikker på, at mine data ikke havner i forkerte hænder?"
Google 2018: Objectives for AI Applications	"We will incorporate our privacy principles in the development and use of our AI technologies. We will give opportunity for notice and consent, encourage architectures with privacy safeguards, and provide appropriate transparency and control over the use of data."
Government of the Netherlands 2019: Strategic Action Plan for Artificial Intelligence	"Privacy may be violated if the processing of personal data does not meet the requirements of fairness and transparency in the GDPR. Privacy issues play a role in areas such as facial recognition technology, big data and techniques in which personal data are derived from other data." (s.41)
Herschel & Miori 2017: Ethics & Big Data	"Employing Social Contract Theory, one can say that an individual has the right to privacy, but also the duty not to invade the privacy of others. That said, Big Data clearly poses a challenge to both. Big data compromises moral rules and duties because in many ways it has rapidly become too powerful, too pervasive, and too essential to day-to-day life." (s.35)
IBM 2019: Everyday Ethics for Artificial Intelligence	"AI must be designed to protect user data and preserve the user's power over access and uses" (s.40) "It is your team's responsibility to keep users empowered with control over their interactions. Pew Research recently found that being in control of our own information is "very important" to 74% of Americans. The European

	Commission found that 71% of EU citizens find it unacceptable for companies to share information about them without their permission. These percentages will rise as AI is further used to either amplify our privacy or undermine it. Your company should be fully compliant with the applicable portions of EU's General Data Protection Regulation and any comparable regulations in other countries, to make sure users understand that AI is working in their best interests" (s.40) Til princippet hører også "Recommended actions to take" og eksempler (se s. 42-44).
Microsoft : Microsoft AI Principles	"AI systems should respect privacy."
Mishra 2020: International Trade Law Meets Data Ethics: A Brave New World	"Privacy and security in the online context have been long recognised in the international policy community as being of utmost importance. This position is not only reflected in the various declarations of the internet multistakeholder declarations, but also reflected in several domestic laws and regulations. Privacy and security issues could be relevant both in the development of data-driven services (for e.g. principle of ethical design), and how individuals are affected while using such technologies, for instance, when subjected to automated decision-making by algorithms. Increasingly, scholars are also arguing that protecting privacy includes protection of 'group privacy' as Big Data analytics permits discrimination against specific groups of people without specifically looking at the personal data of individual members of the group." (s.15)
Mittelstadt & Floridi 2016: The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts	"Anonymisation and privacy were closely linked in the literature, wherein privacy concerns raised by Big Data practices can be addressed merely by removing identifying information. Anonymisation was frequently seen as the minimum requirement necessary to protect data subjects' privacy in aggregating data, despite the possibility of re-identification through cross-referencing with data concerning ethnic background, locational data, other metadata, health records or even small pieces of identified genetic data." "Unsurprisingly, privacy features very frequently in the literature, often in parallel with anonymisation and confidentiality. Commentary pieces often address privacy issues of Big Data, presumably due to the prevalence of the concept in international legislation and related discussions in applied ethics. In the reviewed literature, numerous concerns were described in terms of privacy, some of which relate to alternative concepts such as autonomy or freedom of information. Links are frequently made with confidentiality, understood as ''the duties that accompany the disclosure of non-public information within a fiduciary, professional or contractual relationship''. Others discussed privacy in terms of the 'invasiveness' of Big Data analysis. Invasiveness was connected in particular to analysis of combined data sets, particularly from geolocation and internet-based sources, even when such data is anonymised."
Montreal Declaration Responsible AI : THE DECLARATION	"Personal spaces in which people are not subjected to surveillance or digital evaluation must be protected from the intrusion of AI and data acquisition and archiving systems.

	The intimacy of thoughts and emotions must be strictly protected from AI and DAAS (Data acquisition and archiving systems) uses capable of causing harm, especially uses that impose moral judgments on people or their lifestyle choices. People must always have the right to digital disconnection in their private lives and AI should explicitly offer the option to disconnect at regular intervals, without encouraging people to stay connected. People must have extensive control over information regarding their preferences. AI must not create individual preference profiles to influence the behavior of the individuals without their free and informed consent. DAAS must guarantee data confidentiality and personal profile anonymity. Every person must be able to exercise extensive control over their personal data, especially when it comes to its collection, use, and dissemination. Access to AI and digital services by individuals must not be made conditional on their abandoning control or ownership of their personal data. Individuals should be free to donate their personal data to research organizations in order to contribute to the advancement of knowledge.”
Regulatory and Ethics Working Group of the Global Alliance for Genomics and Health 2014: Framework for Responsible Sharing of Genomic and Health-Related Data	"• Comply with applicable privacy and data protection regulations at every stage of data sharing, and be in a position to provide assurances to citizens that confidentiality and privacy are appropriately protected when data are collected, stored, processed, and exchanged. Privacy and data protection safeguards should be proportionate to the nature and use of the data, whether identifiable, coded or anonymized. • Forego any attempt to re-identify anonymized data unless where expressly authorized by Law." (s.5)
Richards & King 2014: Big Data Ethics	"We typically think about problems of personal information under the rubric of "privacy." But the Big Data Revolution need not signal the "death of privacy." On the contrary, when we think of "privacy" as more than keeping secrets and recognize it instead as the rules we have to govern information flows, big data's real privacy problem comes into focus. We need rules to regulate the flows of data, which means that the collection of personal data should be the beginning of our privacy conversation and not its end." (s.409) "The objective is to provide individuals control over their personal data so that they can weigh the benefits and costs at the time of collection, use, or disclosure. And the most important principles in practice as the law has evolved are notice (the idea that data processors should disclose what they are doing with personal data) and choice (the idea that people should be able to opt-out of uses of their data that they dislike)." (s.412) "Confidentiality is a kind of privacy that is based on trust and reliance on promises in the context of relationships. With the power of big data to make secondary uses of the private information we share in confidence, restoration of trust in the institutions we share with rests not only with privacy but in the recognition that shared private information can remain "confidential." In other words, private digital information that we share with third parties we trust can still be regulated by privacy law." (s.413) "Confidentiality provides the trust necessary to ensure that better sharing takes place under terms that are clear, allowing the benefits of sharing and the protection of privacy at the same time." (s.419)

SAP AI Ethics Steering Committee 2018: SAP's Guiding Principles for Artificial Intelligence	"Data protection and privacy are a corporate requirement and at the core of every product and service. We communicate clearly how, why, where and when customer and anonymized user data is used in our AI software."
THE CENTRE FOR HUMANITARIAN DATA 2020: GUIDANCE NOTE SERIES DATA RESPONSIBILITY IN HUMANITARIAN ACTION : NOTE #4: HUMANITARIAN DATA ETHICS	"Is the data or its use revealing the identity of an individual or group of people?" (s.2)
The Japanese Society for Artificial Intelligence 2017: The Japanese Society for Artificial Intelligence Ethical Guidelines	"Members of the JSAI will respect the privacy of others with regards to their research and development of AI. Members of the JSAI have the duty to treat personal information appropriately and in accordance with relevant laws and regulations."
The Linux Foundation : Data Values and Principles	"Protect the privacy and security of individuals represented in our data."
The Task Force on Artificial Intelligence of the Agency for Digital Italy 2018: White Paper on Artificial Intelligence at the service of citizens	"A further need, closely linked to the previous one, is to protect the data of individuals. PA must design services based on AI able to guarantee efficiency and prompt response, but also protection of citizens' sensitive data. This requirement, strictly connected to the legal context, has some ethical peculiarities concerning the use that PA can make of the data that has come to its knowledge in contexts different from those in which it was collected. Is it ethically sustainable that PA, through the use of data collected for other purposes, takes action based on the new derived information? Is it ethical to use this data to feed predictive systems?" (s.37)
UK Government, Department for Digital, Culture, Media & Sport 2018: Guidance : Data Ethics Framework	"The use of data must be proportionate to the user need. You must use the minimum data necessary to achieve the desired outcome." (s.5)
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	"For the purposes of this Framework, privacy refers to controlling access to information about persons. Privacy is valuable because the ability to control access to information about persons promotes certain core interests that we have as individuals and groups. These are wide-ranging but include identity interests and the promotion of human autonomous decision-making, as well as freedom from potential harms such as discrimination and stigmatisation that may arise from our data being disclosed. This control may be exercised directly by individuals to whom the data pertains, or by designated persons, such as data custodians whose decisions aim to promote those core individual and group interests" (s.245)
Zimmer 2018: Addressing Conceptual Gaps in Big Data	"Paired with informed consent, protecting subject privacy and confidentiality is an essential component of minimizing harm in research contexts, ranging from

Research Ethics: An Application of Contextual Integrity	the exposure of personal or sensitive information, the divulgence of embarrassing or illegal conduct, or the release of data otherwise protected under law. Principles of research ethics dictate that, when appropriate, researchers must take measures to protect the privacy of subjects and to maintain the confidentiality of any data collected or disseminated. Special privacy considerations are triggered when research involves the collection or monitoring of "private information," which has a specific definition in the US federal guidelines: [A]ny information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information that has been provided for specific purposes by an individual and that the individual can reasonably expect will not be made public (for example, a medical record). (45 CFR 46.102[f]) […] Strategies typically include minimizing the private data collected, creating a means to collect data anonymously, removing or obscuring any personal identifiers within the data as soon as reasonable, and using access restrictions and related data security methods to prevent unauthorized access and use of the research data itself." (s.3)
Retfærdighed	
Brakewood & Poldrack 2013: The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data	"Justice, as a Belmont principle, refers to the concept of distributive justice - a balance of benefit and burden. The risks borne by a research population should no result in data, technology, or advances from which they cannot benefit." (s.673)
Datenethik Kommission 2019: Opinion of the Data Ethics Commission	"Observance of the principles of justice by society and its institutions is another fundamental factor that allows us to live together in peace, prosperity, freedom and democracy. Data and technology have placed enormous influence – both economic clout and the societal sway that results from the former – in the hands of a small number of large companies, and this has raised new questions about a fair economic order. The availability of large volumes of data and the digitalisation of processes e.g. in the workplace and the healthcare sector raises other questions relating to equitable access and distributive justice, however, for example in relation to income and the provision of healthcare; these developments may mean that scarce resources can be distributed more fairly, but they may also mean that individual groups of people suffer disadvantage or discrimination." (s.46)
EDPS Ethics Advisory Group 2018: Towards a digital ethics	"The guarantee of justice in any institution is dependent upon a complex and interwoven systems of information management. Political rights are often deeply intertwined with the free flow of impartial information, transparency and accountability. Criminal justice depends critically on information collected and disseminated about the political context. Criminal investigations are linked to the processing of forensic data and questions of appropriateness and admissibility of data. In criminal justice systems, data-driven algorithmic solutions play a privileged role in the tendency towards performance-oriented management of justice systems. This tendency toward technical management of judicial systems impacts the ecosystem of justice in terms of the presumption of innocence, rules of evidence, processes of justification and the

	ability to contest judicial decisions, non-discrimination, and equal access to justice. The new horizon of predictive litigation may render law firms more selective in the cases and the individuals they are willing to represent, encouraging advocates to assess the value of sources of evidence by algorithm instead of by human judgment." (s.19)
Fairfield & Shtein 2014: Big Data, Big Problems : Emerging Issues in the Ethics of Data Science and Journalism	"The principle of justice is described in the Belmont Report as fairness in distribution of benefits (here, the benefits of research). “An injustice occurs when some benefit to which a person is entitled is denied without good reason or when some burden is imposed unduly.”" (s.40)
Mittelstadt & Floridi 2016: The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts	"Interventions and knowledge developed from Big Data, particularly genomic and microbiomic data, may favour populations from whom data is collected, further exacerbating existing gaps in medical practice and knowledge between ‘‘Euro-Americans of middle to upper socio-economic status’’ and others. Even where studied populations are diverse, formal benefit sharing agreements may be required between data subjects and custodians or researchers to ensure data are not taken from one context purely to benefit individuals in another, similar to the issues faced with pharmaceutical research in the third world. As much should be done to facilitate benefit sharing as possible, as Big Data can allow researchers to meet the moral obligation to maximise the value of data collected from research participants without the need for further data collection which places participants at risk."
Willis 2013: Ethics, Big Data, and Analytics: A Model for Application	""Justice emerges when negotiations are without social differentiation." Using big data in analytics might create an ethical dilemma with the Veil of Ignorance, or a method in which impartial judgment is required. Although data can be de-identified, it is often analyzed in terms of inferential statistics that operate under categories such as gender, ethnicity, and income. Aggregated data can provide incredible knowledge to a college administration, but the model's risk is that the framework of categories can artificially impact decisions. Further, aggregated data eliminates outliers, which also can be problematic. Is it possible or beneficial to perform deep analysis while maintaining the Veil of Ignorance over differentiations? Being aware of potential biases that might emerge when pinpointing particular data can help guide meaningful discussions." (s.5-6)
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	"Justice consists in treating individuals and groups fairly and with respect. This includes the fair distribution of benefits and burdens of data activities (collection, storage, use, linkage, and sharing) and attention to issues of equity." (s.245)
Sikkerhed	
Accenture 2016: Universal principles of data ethics : 12 guidelines for developing ethics codes	"The power and peril of data analytics is that data collected today will be useful for unpredictable purposes in the future. Give due consideration to the possibility that less data may result in both better analysis and less risk." (s.8) "Governance practices should be robust, known to all team members and reviewed regularly [...] Data ethics poses organizational challenges that cannot be resolved by familiar compliance regimes alone. Because the regulatory,

	social, and engineering terrains are so unsettled, organizations engaged in data analytics require collaborative, routine and transparent practices for ethical governance." (s.9)
ACM US Public Policy Council 2017: Statement on Algorithmic Transparency and Accountability	"Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public."
Australian Government : AI Ethics Principles	"Throughout their lifecycle, AI systems should reliably operate in accordance with their intended purpose." "This principle aims to ensure that AI systems reliably operate in accordance with their intended purpose throughout their lifecycle. This includes ensuring AI systems are reliable, accurate and reproducible as appropriate. AI systems should not pose unreasonable safety risks, and should adopt safety measures that are proportionate to the magnitude of potential risks. AI systems should be monitored and tested to ensure they continue to meet their intended purpose, and any identified problems should be addressed with ongoing risk management as appropriate. Responsibility should be clearly and appropriately identified, for ensuring that an AI system is robust and safe."
Bertelsmann Stiftung & iRights.Lab 2019: Algo.Rules : Design rules for algorithmic systems	"The function and potential effects of an algorithmic system must be understood." (s.4) "Those who develop, operate and/or make decisions regarding the use of algorithmic systems must have the necessary expertise and appropriate-to-scale understanding of how the technology functions and its potential effects. Sharing individual and institutional knowledge as well as promoting interdisciplinary exchange across task areas are just as crucial as ensuring appropriate skills development. These approaches should be integrated into the education, training and onboarding of new employees. In addition, interdisciplinary exchange should be an ongoing endeavor that remains open to those who are interested and/or affected." (ibid.) "The objectives and expected impact of the use of an algorithmic system must be documented and assessed prior to implementation." (s.4) "The objectives of an algorithmic system must be clearly defined and information regarding its use must be documented. This includes the underlying data and calculation models. Before an algorithmic system is put to use, an impact assessment should be conducted and documented. Particularly in the case of machine-learning systems and in dynamic areas of application that are subject to frequent change, an impact assessment should be repeated at regular intervals. The risk of discrimination and other consequences affecting individuals and the common good must be taken into consideration. The objectives considered, their underlying values and the use of algorithmic systems must be documented." (ibid.) "The security of an algorithmic system must be tested before and during its implementation." (s.5) "The reliability and robustness of an algorithmic system as well as its underlying data with respect to attacks, access and manipulation must be guaranteed. Security must be built into the architecture of the algorithmic system (security by design). The system must be tested in a protected

	environment prior to implementation. Security precautions must be documented." (ibid.) "An algorithmic system must be manageable throughout the lifetime of its use." (s.6) "In order for an algorithmic system to remain adaptable, everyone involved in its development and implementation must maintain joint control over the system. This involves ensuring broad oversight of the entire system, even when tasks are distributed across various departments within an organization and among several individuals. The complexity of a system's operations must never exceed the capacity of human oversight and a person's capacity to make changes to the system. This applies in particular to machine-learning systems. If this manageability cannot be guaranteed, the algorithmic system in question should not be used." (ibid.)
Chinese Expert Group 2019: Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence	"AI systems should continuously improve transparency, explainability, reliability, and controllability, and gradually achieve auditability, supervisability, traceability, and trustworthiness. Pay close attention to the safety/security of AI systems, improve the robustness and tamper-resistance of AI, and form AI security assessment and management capabilities."
DANSK IT's arbejdsgruppe for dataetik 2018: Dataetik : 18 dataetiske anbefalinger fra DANSK IT	"Data er værdifulde [...] Når så virkeligheden indfinder sig, opdager de fleste, at mange af disse vidunderlige data ikke er så anvendelige og værdifulde endda. De viser sig ofte at være forældede, dårligt formaterede og med datamodeller, der passer til et eller andet specifikt behov, som var der på det tidspunkt, hvor modellerne blev lavet. Man opdager som oftest også, at data er dårligt dokumenterede, har uigennemskuelige feltnavne og savner sammenhæng i nøglerne. Man finder ud af, at felter, som var beregnet til ét formål, undervejs er blevet brugt til et eller flere andre formål af brugere, som syntes, det var lettere end at overbevise it-afdelingen eller leverandøren om at ændre i systemet." (s.21) "[Risiko for, at d]ata forstås og fortolkes måske på en anden måde, end de blev i den kontekst, de blev skabt i [...] Det er uundgåeligt, at der vil være et kontekstab og en risiko for fejltolkning, når oplysninger bruges på tværs af tid og sted [...] De medarbejdere, der arbejder med sammensatte data, bør lære, hvordan de omgås dem kritisk og tager højde for, hvor de kommer fra og hvilken sammenhæng, de er skabt i [...] De, der udvikler algoritmer eller sætter parametrene for selvlærende systemer, bør på samme måde forholde sig kritisk til hvilke data, der kan og bør indgå og med hvilken valør. Måske skal data kasseres, fordi valøren er for usikker, eller udtages til manuel tolkning (af kildekritiske medarbejdere)." (s.22-23)
Datenethik Kommission 2019: Opinion of the Data Ethics Commission	"Algorithmic systems also give rise to crucial security questions. The context of use may promote or jeopardise user security. Security is relevant from an ethical and legal perspective because of the role it plays in protecting high-ranking values, such as an individual's physical and mental health and his or her privacy, or public security, peace, and free and equal democratic elections. Security can relate to collecting and using data, which means that the concept also has a bearing on the protection of privacy. The major data scandals that have hit the headlines in recent years have made it clear that privacy breaches and the use of personal data for manipulative purposes can have far-reaching – and sometimes political – consequences." (s.45)

Drew 2016: Data science ethics in government	"It is really important to be honest about the level of confidence in the insight as this will affect the decision made on the back of it, and to keep track of the provenance of the data. Policymakers or operational staff should work together with the data scientist to iterate the model, check that it is working well and be able to explain any unintuitive anomalies." (s.6-7) "We know that the public is justifiably concerned about people's data being lost or stolen. For participants involved in the public research, security was a basic and obvious point. Government has a statutory duty to protect the public's data, and as such it is vital that appropriate security measures are in place." (s.8)
EDPS Ethics Advisory Group 2018: Towards a digital ethics	"The development of human societies has also taken the form of institutionalising trust. As a concept, trust is related to the notions of risk and uncertainty. Trust has grown in importance in the evolution of information technologies as a bridge between technical and moral aspects of technically assisted communication systems. It does however appear prominently wherever the European Commission seeks to advance technological innovation against the apparent or proven resistance of public trust, such as in the Digital Agenda for Europe (2010), the Framework for Building Trust in the Digital Single Market (2011), the Cloud Computing Strategy (2012), the Cybersecurity Strategy (2013) or the much-heralded A Digital Single Market for Europe (2015). Crucially, trust has a double-meaning in data protection. One is a technologically-oriented, functional or knowledge concept: trust in a technology refers to the confidence that it will not fail in its pure functionality, that its design and engineered properties will carry out their expected function. The second, trust is a moral concept referring to belief and re-liance in a person or organisation that they will honour explicit or implicit promises and commitments." (s.20)
EU Commision for the Efficiency of Justice 2018: European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment	"With regard to the processing of judicial decisions and data, use certified sources and intangible data with models elaborated in a multi-disciplinary manner, in a secure tehnological environment. ■Designers of machine learning models should be able to draw widely on the expertise of the relevant justice system professionals (judges, prosecutors, lawyers, etc.) and researchers/lecturers in the fields of law and social sciences (for example, economists, sociologists and philosophers). ■Forming mixed project teams in short design cycles to produce functional models is one of the organisational methods making it possible to capitalise on this multidisciplinary approach. ■Existing ethical safeguards should be constantly shared by these project teams and enhanced using feedback. ■Data based on judicial decisions that is entered into a software which implements a machine learning algorithm should come from certified sources and should not be modified until they have actually been used by the learning mechanism. The whole process must therefore be traceable to ensure that no modification has occurred to alter the content or meaning of the decision being processed. ■The models and algorithms created must also be able to be stored and executed in secure environments, so as to ensure system integrity and intangibility." (s.10)

FAT/ML : Principles for Accountable Algorithms and a Social Impact Statement for Algorithms	“Identify, log, and articulate sources of error and uncertainty throughout the algorithm and its data sources so that expected and worst case implications can be understood and inform mitigation procedures.”
Forsikring & Pension : Cool eller creepy? : Databrug og dataetiske principper i forsikrings- og pensionsbranchen	"Forsikrings- og pensionsbranchen er en kritisk sektor for det danske samfund, og datasikkerhed er en forudsætning for at kunne tale om dataetik. Derfor arbejder branchen kontinuerligt på at have et meget højt sikkerhedsniveau og sikkerhedskultur, således at vi passer bedst muligt på dig og dine data – du skal kunne være sikker på, at dine data ikke havner i forkerte hænder." (s.12)
Future of Life 2017: ASILOMAR AI PRINCIPLES	“AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.”
Google 2018: Objectives for AI Applications	"We will continue to develop and apply strong safety and security practices to avoid unintended results that create risks of harm. We will design our AI systems to be appropriately cautious, and seek to develop them in accordance with best practices in AI safety research. In appropriate cases, we will test AI technologies in constrained environments and monitor their operation after deployment."
Government of Canada: Our guiding principles	"Understand and measure the impact of using AI by developing and sharing tools and approaches."
Government of UK 2019: Understanding artificial intelligence ethics and safety	"The technical sustainability of these systems ultimately depends on their safety, including accuracy, reliability, security, and robustness."
ITI (Information Technology Industry Council) : AI Policy Principles : Executive Summary	"Technologists have a responsibility to ensure the safe design of AI systems. Autnonomous AI agents must treat the safety of users and third parties as a paramount concern, and AI technologies should strive to reduce risks to humans. Furthermore, the development of autonomous AI systems must have safeguards to ensure controllability of the AI system by humans, tailored to the specific context in which a particular system operates." (s.3) "To promote the responsible use of data and ensure its integrity at every stage, industry has a responsibility to understand the parameters and characteristics of the data, to demonstrate the recognition of potentially harmful bias, and to test for potential bias before and throughout the deployment of AI systems. AI systems need to leverage large datasets, and the availability of robust and representative data for building and improving AI and machine learning systems is of utmost importance." (s.3)
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	“Designers and users of AI systems must remain aware that these technologies have transformative effects on individuals and society. They must thereby proceed with a continuous sensitivity to real-world impacts. They must also keep in mind that the technical sustainability of these systems depends on their safety: their accuracy, reliability, security, and robustness.” (s.12)

Microsoft : Microsoft AI Principles	“AI systems should be secure.” “AI systems should perform safely.” “AI systems should perform reliably.”
Montreal Declaration Responsible AI : THE DECLARATION	“It is necessary to develop mechanisms that consider the potential for the double use - beneficial and harmful - of AI research and AI development (whether public or private) in order to limit harmful uses. When the misuse of an AI endangers public health or safety and has a high probability of occurrence, it is prudent to restrict open access and public dissemination to its algorithm. Before being placed on the market and whether they are offered for charge or for free, AI must meet strict reliability, security, and integrity requirements and be subjected to tests that do not put people's lives in danger, harm their quality of life, or negatively impact their reputation or psychological integrity. These tests must be open to the relevant public authorities and stakeholders. The development of AI must pre-empt the risks of user data misuse and protect the integrity and confidentiality of personal data. The errors and flaws discovered in AI and SAAD should be publicly shared, on a global scale, by public institutions and businesses in sectors that pose a significant danger to personal integrity and social organization.”
Norwegian Ministry of Local Government and Modernisation 2020: National Strategy for Artificial Intelligence	“AI must be built on technically robust systems that prevent harm and ensure that the systems behave as intended. The risk of unintentional and unexpected harm must be minimised. Technical robustness is also important for a system's accuracy, reliability and reproducibility.” “AI built on personal data or on data that affects humans must respect the data protection regulations and the data protection principles in the GDPR.”
OECD 2019: Recommendation of the Council on Artificial Intelligence	“AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk. To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art. AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.”
Regulatory and Ethics Working Group of the Global Alliance for Genomics and Health 2014: Framework for Responsible Sharing of Genomic and Health-Related Data	"• Store and process the data collected, used and transferred in a way that is accurate, verifiable, unbiased, proportionate, and current, so as to enhance their interoperability and replicability and also preserve their long-term searchability and integrity. • Ensure feedback mechanisms on the utility, quality, security, and accuracy of data, and their annotations, with a view to improving quality and interoperability and appropriate re-use by others. • Establish proportionate data security measures that mitigate the risk of unauthorized access, data loss and misuse.

	• Understand the issues related to lawful requests for data based on law enforcement, public health, or national security concerns." (s.4-5) "• Ensure, where appropriate, the sustainability of the data generated for future use, through both archiving and using appropriate identification and retrieval systems, and through critical appraisal of the mechanisms and systems used for sharing genomic and health-related data." (s.6) "• Dedicate education and training resources so as to advance data sharing and data management and to constantly improve data quality and integrity. Education and training resources should also be dedicated to: fostering and maintaining good records about the effects and impact of data sharing; raising awareness about national health priorities and distribution of health services; building capacity and data sharing infrastructure in countries; and, working towards the building of an evidence base about the advantages and potential limitations of data sharing." (s.6)
SAP AI Ethics Steering Committee 2018: SAP's Guiding Principles for Artificial Intelligence	"As with any of our products, our AI software is subject to our quality assurance process, which we continuously adapt when necessary. Our AI software undergoes thorough testing under real-world scenarios to firmly validate they are fit for purpose ad that the product specifications are met. We work closely with our customers and users to uphold and further improve our systems' quality, safety, reliability and security."
Smart Dubai : ARTIFICIAL INTELLIGENCE PRINCIPLES & ETHICS	"AI systems should be safe and secure, and should serve and protect humanity."
THE CENTRE FOR HUMANITARIAN DATA 2020: GUIDANCE NOTE SERIES DATA RESPONSIBILITY IN HUMANITARIAN ACTION : NOTE #4: HUMANITARIAN DATA ETHICS	"Is the model (or underlying data) codifying the current state of the world and thereby making it harder to change? (e.g. are we building models that perpetuate or encore [sic] past mistakes)" (s.2)
The Japanese Society for Artificial Intelligence 2017: The Japanese Society for Artificial Intelligence Ethical Guidelines	"Members of the JSAI must verify the performance and resulting impact of AI technologies they have researched and developed. In the event that potential danger is identified, a warning must be effectively communicated to all of society. Members of the JSAI will understand that their research and development can be used against their knowledge for the purposes of harming others, and will put in efforts to prevent such misuse. If misuse of AI is discovered and reported, there shall be no loss suffered by those who discover and report the misuse."
The Public Voice 2018: Universal Guidelines for Artificial Intelligence	"An AI system should be deployed only after an adequate evaluation of its purpose and objectives, its benefits, as well as its risks." "Institutions must assess the public safety risks that arise from the deployment of AI systems that direct or control physical devices, and implement safety controls." "Institutions must ensure accuracy, reliabillity, and validity of decisions." "Institutions must secure AI systems against cybersecurity threats."

	“Institutions must establish data provenance and assure quality and relevance for the data input into algorithms.”
The Task Force on Artificial Intelligence of the Agency for Digital Italy 2018: White Paper on Artificial Intelligence at the service of citizens	"Machine learning systems need data which is “annotated” by human beings (supervised learning) or at least selected and prepared (unsupervised learning). This also includes errors or bias introduced, even inadvertently, by the designers, replicating them in all future applications. For example, datasets with bias they propagate the same evaluation errors in the meaning of an image or a concept, as happened, for example, with certain algorithms used to prevent crimes, in which the data was compromised by a historical series that emphasised ethnic differences. Or unbalanced datasets, that overestimate or underestimate the weight of certain variables in the reconstruction of the cause-effect relationship necessary to explain certain events and, above all, to predict them." (s.36)
UK Government, Department for Digital, Culture, Media & Sport 2018: Guidance : Data Ethics Framework	"You must have an understanding of the relevant laws and codes of practice that relate to the use of data. When in doubt, you must consult relevant experts." (s.5) "Insights from new technology are only as good as the data and practices used to create them. You must work within your skillset recognising where you do not have the skills or experience to use a particular approach or tool to a high standard." (s.5) "Data used to inform policy and service design in government must be well understood. It is essential to consider the limitations of data when assessing if it is appropriate to use it for a user need." (s.5)
UK Statistics Authority : Data Ethics	“The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.”
Velfærd	
Amsterdam Economic Board (TADA): The 6 principles of our manifesto	"Data and technology should contribute to the freedom of people. Data are meant to serve the people. To be used as seen fit by people to benefit their lives, to gather information, develop knowledge, find room to organise themselves. People stay in control over their data."
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	"• Design and deploy AI systems to foster and to cultivate the welfare of all stakeholders whose interests are affected by their use • Do no harm with these technologies and minimise the risks of their misuse or abuse • Prioritise the safety and the mental and physical integrity of people when scanning horizons of technological possibility and when conceiving of and deploying AI applications." (s.10-11)
Montreal Declaration Responsible AI : THE DECLARATION	“AI must help individuals improve their living conditions, their health, and their working conditions. AI must allow individuals to pursue their preferences, so long as they do not cause harm to other sentient beings. AI must allow people to exercise their mental and physical capacities. AI must not become a source of ill-being, unless it allows us to achieve a superior well-being than what one

	could attain otherwise. AI use should not contribute to increasing stress, anxiety, or a sense of being harassed by one's digital environment. AI must not threaten the preservation of fulfilling moral and emotional human relationships, and should be developed with the goal of fostering these relationships and reducing people's vulnerability and isolation. AI must be developed with the goal of collaborating with humans on complex tasks and should foster collaborative work between humans AI should not be implemented to replace people in duties that require quality human relationships, but should be developed to facilitate these relationships. Health care systems that uses AI must take into consideration the importance of a patient's relationships with family and health care staff.”
The Japanese Society for Artificial Intelligence 2017.: The Japanese Society for Artificial Intelligence Ethical Guidelines	“Members of the JSAI will contribute to the peace, safety, welfare, and public interest of humanity. They will protect basic human rights and will respect cultural diversity. As specialists, members of the JSAI need to eliminate the threat to human safety whilst designing, developing and using AI.”
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	“AI should be developed to enhance the quality of life.”
Willis 2013: Ethics, Big Data, and Analytics: A Model for Application	"Seek the greatest happiness for the greatest number." College administration ought to consider what utilitarian "happiness" is in terms of student success. Big data can provide unique insight into what we might call "happiness," provided it is framed in terms of learning objectives and grade outcomes. Both of these are quantifiable and, thus, demonstrable of what really works; big data can provide the matrix of successful behaviors that can contribute to outcomes indicative of real learning." (s.5)
Værdighed	
Danish Expert Group on Data ethics 2018: Data for the Benefit of the People : Recommendations from the Danish Expert Group on Data Ethics / Data i menneskets tjeneste : Anbefalinger fra Ekspertgruppen om dataetik	"Human dignity outweighs profit." (s.8) / "Menneskets iboende værdighed skal vægtes over profit." (s.8)
Datenethik Kommission 2019: Opinion of the Data Ethics Commission	"Human dignity, which from an ethical viewpoint is synonymous with the unconditional value of every human being and which is enshrined as a “fundamental constitutional principle” in the constitutional order, is of foundational and supreme importance. It follows from the principle of human dignity that every individual merits respect, regardless of his or her attributes and achievements. Protecting the value which is inherent in every human being and which does not need to be acquired also implies that human beings are not ranked in a classifying system across various spheres of life and

	activities (“super scoring”) or labelled like an object with a price and treated accordingly. The fact that each human is an individual rather than a pattern made up of data points must also be borne in mind at all times in situations where human behaviour is measured and these measurements are processed by algorithmic systems. Algorithmic systems must therefore always be designed in such a way that they can cater to each human’s claim to individuality." (s.43)
Den danske regering 2019: Dansk National Strategi for Kunstig Intelligens	"Menneskets værdighed skal respekteres i udvikling og anvendelse af kunstig intelligens. Kunstig intelligens må ikke gøre skade på mennesker og skal understøtte retssikkerhed og ikke uberettiget stille personer dårligere. Kunstig intelligens skal respektere demokratiet og demokratiske processer, og det må ikke anvendes til at krænke menneskets grundlæggende rettigheder." (s.28)
EDPS Ethics Advisory Group 2018: Towards a digital ethics	"The notion of human dignity in the European intellectual tradition has its origins in the Kantian idea that human beings are to be understood as ends in themselves and never as a means alone. Since the end of World War II in particular, this concept has had a key impact on International Human Rights Law, in legal scholarship and in jurisprudence. It has also been acknowledged as a foundational value in most human rights instruments. The Universal Declaration of Human Rights of 1948 recognises that the inherent dignity and the equal and inalienable rights of all members of the human family are the foundation of freedom, justice and peace in the world. It appears in every iteration of the Treaty of European Union (or Community) beginning from the Treaty of Rome (1957) together with freedom, democracy, equality, rule of law, and respect for human rights, as one of the core values of the European project. The Charter of Fundamental Rights of the European Union explicitly acknowledges the foundational role of the value of human dignity." (s.16)

Bilag: Litteraturstudie 3 (dataetiske dilemmaer)

Litteraturstudiet er foretaget ved gennemsyn af kilder fra litteraturstudie 1 (definitioner af dataetik) og 2 (dataetiske principper og værdier) og desuden ved søgninger på scholar.google.com, google.com, infomedia.dk og forskningsdatabasen.dk i oktober og november 2020. Litteraturstudiet benyttede følgende søgeord på hver af de fire databaser:

- Data etik dilemma
- Datatetik dilemma
- Data dilemma
- Data problem
- Data udfordring
- AI dilemma
- AI ethics dilemma
- AI ethics cases
- Data ethics cases
- Data cases
- Data issues
- Kunstig intelligens dilemma
- Data ethics dilemma
- Data ethics dilemmas
- Data challenges
- AI challenges
- Dataetik eksempel
- Data eksempel
- Dataetik eksempler
- Data eksempler

Søgningerne genererede forventeligt mange hundrede resultater. Disse resultater er filtreret to gange. Ved de umiddelbare søgninger frasorteredes indlysende irrelevante resultater. De tilbageværende resultater blev indhentet og studeret. Ved læsning af disse resultater blev yderligere en håndfuld resultater sorteret fra som irrelevante.

Litteraturstudiet har identificeret 152 relevante cases, hvoraf 46 stammer fra akademiske kilder, 43 stammer fra nyhedsmedier, mens 63 stammer fra policy-dokumenter og øvrige kilder. Kilderne optræder nedenfor citeret med forfatter, årstal, titel og hyperlink.

Kilde(r)	Berørte værdier/ principper	Uddrag
Politiken 2019: Lektor er forarget: Vores børns Aula-data bør ikke gemmes hos Amazon Ekstrabladet 2019: Dine børns fortrolige data ligger hos verdens rigeste mand Computerworld 2019: Danske forældre	Sikkerhed	"Hvordan bliver de mange oplysninger om mit barns trivsel og faglighed opbevaret sikkert og forsvarligt i udlandet? Det spørger danske forældre i almindelighed og organisationen Skole og Forældre i særdeleshed sig selv om i forbindelse med, at langt de fleste af landets skoler efter efterårsferien skifter Skoleintra og dermed Elevintra samt Forældreintra ud med et nyt IT-system kaldet Aula [...] Amazon Web Services (AWS) skal stå for driften. AWS er den amerikanske IT-gigants hurtigstvoksende forretningsområde. Det betyder, at de mange data vil blive opbevaret uden for Danmarks grænser i et eller flere EU-lande. Data bliver opbevaret udenlands, fordi AWS kan levere den bedste og mest sikre løsning, lyder det fra kommunernes IT-fællesskab, Kombit, som står bag Aula. Af hensyn til sikkerheden vil Kombit ikke oplyse, i hvilke lande data bliver opbevaret [...] Opbevaringen i udlandet skaber usikkerhed blandt forældre, lyder det fra formanden for organisationen Skole og Forældre, der er

bekymrede over Amazon-aftale - men Amazon garanterer for datasikkerhed: "Der er ingen gråzoner. Vi lever op til reglerne" Berlingske 2019: Oplysninger om dit barns karakterer og trivsel bliver gemt i et hemmeligt land		landsorganisation for skolebestyrelser og forældre til børn i folkeskolen. »I vores hovedbestyrelse, hvor der sidder forældre fra skoler i hele landet, har vi diskuteret sikkerheden flere gange. Jeg hører det også fra forældre, når jeg møder dem, og jeg læser bekymringer fra forældre på de sociale medier,« siger formand Rasmus Edelberg: »Vi føler os usikre, fordi vi ikke ved nok om det. Vi ved, at man kan opbevare data i udlandet på en sikker måde, men vi ved ikke, hvordan man opbevarer data i forbindelse med Aula, og det er et problem.« Det er bl.a. spørgsmål som, hvordan det konkret sikres, at data ikke ender i de forkerte hænder, og hvordan man arbejder for at undgå hackerangreb, der ifølge formanden skaber usikkerhed" (Berlingske 2019)
Bruin & Floridi 2016: *The Ethics of Cloud Computing*	Sikkerhed; Privatliv; Gennemsigtighed	"Businesses and individual users alike are embracing online software in order to process, share and synchronize data, recruit personnel, organize customer services and sales, and for an increasing number of other purposes. Computing resources (especially software, memory space, CPU power, and maintenance routines) are becoming services on demand, offered by online providers that store and process files in large datacentres. This new Information Technology (IT) paradigm of cloud computing offers huge advantages in terms of installation, configuration, updating, compatibility, costs and computational power (Zhang et al. 2010), and in the last few years cloud computing has already provided enormous benefits to a large number of users. However, it also comes with a number of potential risks. The year 2010, for instance, witnessed a huge cyber attack on the popular cloud email services of Gmail, and the sudden discontinuation of cloud services to WikiLeaks by Amazon. There followed the 2013 NSA spying scandal, the 2014 nude photo iCloud hack and the Sony hack, with hackers increasingly turning to the cloud" (p. 22) "[...] many clouders are unaware of what cloud computing really amounts to. We have argued that this is due to a lack of interlucent

		communication between the cloud computing industry and its customers, and showed that competing with integrity in this emerging market requires of hosting companies and cloud services providers that they do their utmost to ensure that customers understand what they buy. Second, we defended the claim that regulation of the hosting companies and the cloud services providers should be at a minimum, because proscriptive pressure here risks slowing down innovation. Yet regulation of the business customers of the cloud services providers is urgently needed. Hosting companies and cloud services providers move their customers' data with high frequency from one datacentre to another so as to enable efficient use of storage space. This is one of the innovations that marks cloud computing. But it is currently unsuitable, we have argued, to store lawyers' client data, for instance, or sensitive military, business or medical data. Disagreeing with several professional associations, we defended, for instance, the claim that lawyers should be forbidden to store client data in the cloud" (p. 37)
Amnesty 2019: AMNESTY: STATENS ULOVLIGE MASSEOVERVÅGNING SKAL BREMSES Magisterbladet 2019: Fagforeninger kræver stop for teleselskabers logning af data: "Vi har ikke brug for masseovervågning af alle danskere"	Privatliv	"Regeringen vil fortsætte flere års ulovlig praksis med at overvåge danskernes telefonsamtaler og færden på internettet. Amnesty opfordrer Folketingets Retsudvalg til at skride ind og prioritere borgernes retssikkerhed" "Skandale omkring teleselskabernes logning af danskernes data risikerer at sætte sammenhængskraften i vores samfund over styr og skal stoppe nu. Den danske regering skal rette ind og følge EU-lovgivningen på området, kritiserer fagforeninger" (magisterbladet.dk)

Schermer 2011: *The limits of privacy in automated profiling and data mining*	Lighed	"Automated profiling of groups and individuals is a common practice in our informationsociety. The increasing possibilities of data mining significantly enhance the abilities to carry out such profiling. Depending on its application, profiling and data mining may cause particular risks such as discrimination, de-individualisation and information asymmetries.In this article we provide an overview of the risks associated with data mining and the strategies that have been proposed over the years to mitigate these risks. From there we shall examine whether current safeguards that are mainly based on privacy and dataprotection law (such as data minimisation and data exclusion) are sufficient"
EESC 2016: The ethics of Big Data: Balancing economic benefits and ethical questions of Big Data in the EU policy context	Etik	"A more complicated matter revolves around how to consider a user's data that was produced after processing the original dataset: are they still a user's data, or do they belong to the company that carried out the analyses? Or to the company that collected the original data? This issue has been tackled by limiting the place where physical storage of data takes place, namely the countries where servers are based, and the EU approach has been to progressively restrict the possibility for EU citizens' data to be stored out of the so-called "Euro cloud". This approach still leaves the problem of where already processed data is stored, and does not resolve the ethical dilemma of how data ownership is defined philosophically, before passing to a more down-to-earth approach of law and policy making" (p. 57)
EESC 2016: The ethics of Big Data: Balancing economic benefits and ethical questions of Big Data in the EU policy context	Sikkerhed; Autonomi; Privatliv	"The ubiquitous use of CCTV circuits, coupled with the GPS positioning capabilities built in mobile devices, and the use of credit cards and ATM cards for payments and withdrawals represent only some of the available means to track one's position over time. This ease of tracking has surely increased safety (or the perception of it) across Europe and allows for a more effective and focused workflow of police forces during investigations, but this might come at a cost. Active surveillance is an extremely effective mean

		of limiting citizens' liberties, and has already been used as such by totalitarian government over the course of EU's history. When surveillance is carried out by employers (e.g. by controlling or restricting access to websites, by encouraging the use of a company's devices, by installing cameras in the workplace) this might lead to an increased stress level of workers, which often translates into lower productivity. The awareness of the possibility of being watched at any moment, furthermore, creates an ideal panopticon in which an individual's actions tend to conform to the expected norm, as shown by field experiments102" (p. 57-58)
Dagbladet Roskilde 2019: Dine data skal være dine - og dine alene	Demokrati; Autonomi	"Men en ting er, hvordan store virksomheder kan bruge data til at tjene penge. Noget andet er, hvordan staten kan bruge data til at styre dig. Og her er skrækeksemplet Kina, mener Pernille Tranberg. - Kina er det, jeg kalder for et »datadiktatur«, hvor staten styrer borgerne gennem data. Det er noget af det, de unge i Hongkong kæmper imod, siger hun. I Kina er ting som ansigtsgenkendelse og kontrol gennem data langt mere udbredt end i resten af verden, og det udnytter staten fuldt ud. Blandt andet er de kinesiske myndigheder ved at rulle et datadrevent pointsystem ud blandt landets over en milliard indbyggere. - En science fiction-film kunne ikke fantasere sig til det, men det bliver foldet ud i Kina i 2020, og millioner lever allerede under det, fortæller Pernille Tranberg. Essensen i systemet er, at man fra fødslen får en score, altså et antal point. Du kan tjene eller tabe point, alt efter hvordan du opfører dig. Og det er myndighederne, der bestemmer, hvornår og hvordan du skal straffes. - Går du over for rødt lys, så ryger du ned i point. Opfører du dig ordentligt, så ryger du op. Og den kinesiske regering definerer, hvad der er rigtigt og forkert. Og så er det afhængigt af din score, hvor dine børn får lov til at gå i skole, og om du skal tage bumletog eller fly, når du skal besøge familien i den anden ende af Kina, forklarer Pernille Tranberg [...] I Europa ønsker vi ikke et datadiktatur [...] Europa, der på grund af blandt andet EU's datalovgivning, kendt som GDPR, er det, Pernille Tranberg kalder et »datademokrati«. - GDPR-lovgivningen giver os nogle

		basale rettigheder over vores egne data. Du kan bede om dine data hos Facebook og få dem slettet, det er din ret. Du kan få dem overført og bruge dem i en anden tjeneste, på din egen cloud (dit private netværk, red.), hvis du vil, siger hun og peger på, at data har det bedst, når det primært er dig, der har kontrollen over dem. - Så kan man selv berige sine egne data i forskellige tjenester og få en bedre økonomi, et bedre helbred, bedre ruter til arbejde og så videre og så videre. Men man skal have databeskyttelsen først, og kontrollen skal ikke ligge hos Facebook eller Google eller staten. Det skal ligge hos dig. Ellers opnår vi ikke datademokrati, fastslår Pernille Tranberg" (Sjællandske Slagelse 2020)
Dagbladet Roskilde 2019: Dine data skal være dine - og dine alene The New York Times 2019: Apple Card Investigated After Gender Discrimination Complaints BBC News 2019: Apple's 'sexist' credit card investigated by US regulator **CNN 2019: Apple Card is accused**	Lighed	"- Mennesket skal have gavn af det her, og hvis vi ikke har det, men det primært er for at tjene penge eller spare penge, er det ikke dataetisk, understreger hun og påpeger, at kunstig intelligens på baggrund af data også kan ende med at være diskriminerende og giver et eksempel fra USA. Her fik en mand et kreditkort og en høj kreditværdighed. Hans kone, der tjente samme løn, fik en langt lavere kreditværdighed. - Det er så, fordi nogle algoritmer - bevidst eller ubevidst - diskriminerer mod kvinder, forklarer Pernille Tranberg [...]" (Sjællandske Slagelse 2020) "The Bloomberg news agency reported on Saturday that tech entrepreneur David Heinemeier Hansson had complained that the Apple Card gave him 20 times the credit limit that his wife got. In a tweet, Mr Hansson said the disparity was despite his wife having a better credit score. Later, Mr Wozniak, who founded Apple with Steve Jobs, tweeted that the same thing happened to him and his wife despite their having no separate bank accounts or separate assets [...] Mr Hansson said in a tweet: "Apple Card is a sexist program. It does not matter what the intent of individual Apple reps are, it matters what THE ALGORITHM they've placed their

of gender bias. Here's how that can happen		complete faith in does. And what it does is discriminate" (BBC November 2019) "Apple co-founder Steve Wozniak says Apple Card discriminated against his wife [...] The New York Department of Financial Services is looking into allegations of gender discrimination against users of the Apple Card, which is administered by Goldman Sachs. The allegations blew up on Twitter Saturday after tech entrepreneur David Heinmeier Hansson wrote that Apple Card offered him twenty times the credit limit as his wife, although they have shared assets and she has a higher credit score. Many other users voiced similar experiences — including Apple (AAPL) co-founder Steve Wozniak. Wozniak said his credit limit was 10 times that of his wife, despite the fact that they share all assets and accounts. "Some say the blame is on Goldman Sachs, but the way Apple is attached, they should share responsibility," Wozniak tweeted" (CNN Business 2019)
Jydske Vestkysten 2020: Kommune tvang forsyningsselskab til at forære dyre data om Als-Fyn bro væk	Etik	"Det endte i priskrig, da konsulenthuset rambøll sidste år anmodede Sønderborg Forsyning om at få udleveret data til en rapport om Als-Fyn broen. Forsyningen krævede en halv million, men rambøll ville kun betale et par tusinde kroner. Kommunen skar igennem og tvang forsyningen til at forære dataene væk trods stærk modstand fra forsyningens direktør [...] Sagen tager sin begyndelse, da konsulentvirksomheden Rambøll i december 2018 henvender sig til Sønderborg Forsyning for at få udleveret en række data om havbunden. De skal bruges til en rapport, som Rambøll udarbejder for Vejdirektoratet om Als-Fyn forbindelsen. Rambøll bliver mødt med en pris på en halv million kroner for dataene. Det er ifølge forsyningsselskabet en rimelig pris, idet undersøgelser og de tilhørende data oprindeligt har kostet fem millioner kroner, viser korrespondance, som JydskeVestkysten har fået aktindsigt i. Rambøll vil dog kun give et symbolsk beløb på et

		par tusinde kroner, hvilket ifølge virksomheden er normal praksis for udveksling af geotekniske oplysninger mellem bygherrer og firmaer" (Jydske Vestkysten 2020)
Sjællandske Nyheder 2020: Fodboldspillere ønsker magt over egne data The Athletic 2020: There is a value on information. It is no different for footballers	Autonomi	"Bettingselskaber og spilfirmaer bruger meget af den data, fodboldklubberne indsamler. Men der er tvivl om ejerskabet af de samme data [...] Liverpools offensive stjerne Mohamed Salah scorede i den forgange Premier League-sæson 19 mål. Det gjorde han på 132 forsøg, hvoraf 59 ramte mål. Undervejs tilbagelagde den 28-årige egypter 314,8 kilometer og udførte 413 spurter med en topfart på 34,93 kilometer i timen [...] det er oplysninger som disse, som en lang række fodboldspillere nu ønsker at få kontrol over gennem et storstilet sagsanlæg. Det skriver sportssiden The Athletic. Ifølge The Athletic har over 400 unavngivne spillere fra de bedste britiske fodboldrækker tilsluttet sig sagen, der har fået tilnavnet »Project Red Card«. Årsagen er, at eksempelvis bettingselskaber og spilfirmaer i stigende grad lukrerer på den voksende mængde data, der forlader grønsværen" (Sjællandske Nyheder, August 2020)
Berlingske 2020: Myndighederne ønsker adgang til mobildata – eksperter tager situationen med ro	Godgørenhed; Velfærd; Privatliv; Autonomi	"Myndighederne ønsker at få adgang til oplysninger fra de danske mobilnet for at kortlægge, hvor borgerne befinder sig, så myndighederne bedre kan sætte ind mod spredning af coronavirussen. Ønsket følges nøje, men får ikke umiddelbart eksperter og andre til at trykke på den røde alarmknap. Teleindustrien, som er brancheorganisation for de danske tele- og internetudbydere, modtog 19. marts en henvendelse fra Statens Serum Institut om at få adgang til teleoplysninger, som kan hjælpe i kampen mod coronavirus. »Vi er nu i dialog med myndighederne og Rigspolitiet om, hvad de præcis ønsker at få udleveret. Som udgangspunkt ønsker vi selvfølgelig at assistere, hvor vi kan, hvis det giver værdi for myndighederne i bekæmpelse af sygdommen. Vi skal dog sikre os, at det sker på et solidt og juridisk forsvarligt

		grundlag med en klar hjemmel,« forklarer Jakob Willer, direktør i Teleindustrien" (Berlingske Business, marts 2020)
DR Nyheder 2020: Kommune viser detaljeret kort over, hvor de coronasmittede bor: Borgere frygter mistænkeliggørelse	Godgørenhed; privatliv; fairness	"Kommune viser detaljeret kort over, hvor de coronasmittede bor: Borgere frygter mistænkeliggørelse Målet er at vise, at smitten er bredt ude, og ingen kan slappe af, siger kommunaldirektør [...] I Etisk Råd har medlemmerne faktisk netop diskuteret, hvordan man bør holde tungen lige i munden, når man bruger danskernes mange sundhedsdata i bekæmpelsen af covid-19. Og her rammer Brøndby Kommune ved siden af skiven. Det siger medlem af etisk råd og næstformand i SF Lise Müller. - Der er en risiko for, at man stigmatiserer et bestemt område eller en bestemt befolkningsgruppe. Og hvilken værdi er der i, at man udpeger, hvor de bor henne? Kortet er nemlig opgjort efter, hvor de coronasmittede har adresse henne og ikke hvor, de er blevet smittet" (DR oktober 2020)
Skive Folkeblad 2019: Indsamler og sælger brugernes data: Etisk Råd har set sig sure på sundhedsapps Det Etiske Råd 2019: Etik og sundhedsdata En udtalelse fra Det Etiske Råd Det Etiske Råd 2019:	Privatliv; Sikkerhed; Lighed	"Etisk Råd går med en ny rapport i flæsket på de mobilapps og enheder, som smart-watches og fitnessarmbånd, der indsamler data om vores sundhedsvaner. Produkterne bliver samlet set kaldt sundheds-wearables. Problemet er ifølge rådet, at virksomhederne bag disse sundheds-wearables systematisk indsamler brugernes data og sælger dem videre, uden at forbrugerne er klar over det. Det står godt nok nævnt i det alenlange betingelser for at bruge den pågældende app eller enhed, men de fleste trykker bare » accepter « uden at læse betingelserne, påpeger rådet [...] Etisk Råd peger på, at sådanne profiler kan bruges til at skræddersy forsikringstilbud, vælge de sundeste medarbejdere eller frasortere jobansøgere. Som et konkret eksempel nævner Lise von Seelen en nylig sag, hvor en ung kvinde ønskede at optage et lån i sin bank: - Hun havde tidligere i sit liv haft en depression, og derfor krævede

Redegørelse om sundhedswearables og big data		banken som sikkerhed for lånet, at hun tegnede en livsforsikring." (Skive Folkeblad oktober 2019) "De etiske problemer opstår, når myndigheder eller firmaer kan få adgang til data, som gør dem i stand til at gribe ind i vores liv på begrænsende måder, fastslår Rådet. Hvis de kan anvende data til at begrænse vores privatliv og frihed eller til overvågning eller diskriminering, så bør der sættes ind med rammesætning. Rådet kommer med anbefalinger på 4 områder: 1. Om sundhedsvæsenet bør anvende apps udviklet af private firmaer 2. Om der, i forhold til private firmaer, bør eksistere et alternativ til at betale med sine data 3. Om sundhedsvæsenet bør kunne anvende algoritmer til at køre borgernes data opsamlet med wearables sammen med data fra forskellige forvaltninger mhp forebyggende tiltag? 4. Forsikringsselskabers og arbejdsgiveres anvendelse af datagenererede sundhedsdata" (se Rådets anbefalinger her: https://www.etiskraad.dk/~/media/Etisk-Raad/Etiske-Temaer/Sundhedsdata/Publikationer/Udtalelse-om-etik-og-sundhedsdata-2019.pdf)
Ekbia et. al. (2014): *ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review*	Sikkerhed	"British researchers, for instance, recently unveiled a program called Emotive “to map the mood of the nation” using data from Twitter and other social media (BBC, 2013). By analyzing 2,000 tweets per second and rating them for expressions of one of eight human emotions (anger, disgust, fear, happiness, sadness, surprise, shame, and confusion), the researchers claim that they can “help

		calm civil unrest and identify early threats to public safety" (BBC, 2013, para. 3). While the value of a project such as this may be clear from the perspective of law enforcement, it is not difficult to imagine scenarios where this kind of threat identification might function erratically, leading to more chaos than order. Even if we assume that human emotions can be meaningfully reduced to eight basic categories (what of complex emotions such as grief, annoyance, contentment, etc.?), and also assume cross-contextual consistency in the expression of emotions (how does one differentiate the "happiness" of the fans of Manchester United after a winning game from the expression of the "same" emotion by the admirers of the Royal family on the occasion of the birth of the heir to the throne?), technical limitations are quite likely to undermine the predictive value of the proposed system. Available data from Twitter have been shown to have a very limited character in terms of scale (spatial and temporal), scope, and quality (boyd & Crawford, 2012), making some observers wonder if this kind of "nowcasting" could blind us to the more hidden long-term undercurrents of social change (Khan, 2012)" (p. 1529-1530)
Ekbia et. al. (2014): *ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review*	Privatliv	"Today, Big Data presents a new set of privacy concerns in diverse areas such as health, government, intelligence, and consumer data. One new privacy challenge stems from the use of Big Data to predict consumer behavior. A frequently (over)used example is that of the department store Target predicting the pregnancy of expectant mothers in order to target them with coupons at what is perceived to be the optimal time (Duhigg, 2012). In a similar vein, advertisers have recently started assembling shopping profiles of individuals based on compilations of publicly available metadata (such as the geographic locations of social media posts; Mattioli, forthcoming), which has led some commentators to criticize such practices as privacy violations (Tene & Polonetsky, 2012)" (p. 1536)

Ekbia et. al. (2014): *ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review*	Privatliv	"Similar concerns have been expressed in the realm of medical research, where the electronic storage and distribution of individual health data can potentially reveal information not only about the individual but about others related to them. This might happen, for instance, with the public availability of genomic data, as in a recent case that raised objections from the family members of a deceased woman. The subsequent removal of that information marked a legal triumph for the family, but was a worrisome sign for advocates of open access that privacy concerns might significantly slow the progress of Big Data research (Zimmer, 2013). As leading commentators have observed, "[t]hese types of harms do not necessarily fall within the conventional invasion of privacy boundaries," and thus raise new questions for policymakers (Crawford & Schultz, 2014, p. 93)" (p. 1536)
Ekbia et. al. (2014): *ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review*	Privatliv	"Another set of privacy concerns stems from government uses of Big Data. The recent revelation in the U.S. about the NSA's monitoring of e-mail and mobile communications of millions of Americans might just be the tip of a much bigger legal iceberg that will affect the whole system, including the Supreme Court (Risen, 2013; Risen & Lichtblau, 2013). Although the public seems to be divided in its perception of the privacy violations (Shane, 2013), the official defense by the government tends to highlight the point that data gathering did not examine the substance of e-mails and phone calls, but rather focused on more general metadata (Savage & Shear, 2013). Similar themes were raised in a 2012 Supreme Court ruling on the constitutionality of the use of GPS devices by police officers to track criminal suspects (U.S. v. Jones, 2012). Public discussions on approaches to and standards of privacy are complicated by the fact that government officials seem to be concerned less with what they actually do to people's data than with public perceptions of what they do or might do (Nash, 2013)." (p. 1536)

Willis 2013: *Ethics, Big Data, and Analytics: A Model for Application*	Etik	"One of the first full-scale analytics projects, Purdue University's Signals predicts which students are at risk of doing poorly using data such as student demographics and academic history, engagement with online resources, and current performance in a given course.7 Based on the prediction, students receive a red, yellow, or green light; the red light indicates "stop and get help," yellow is "caution you are falling behind," and green is "keep on going." With each light, students receive instructor feedback on their performance and where to go next. Purdue University's work has led to significant gains in overall retention. For example, the four-year retention rate for Purdue students in the 2007 cohort without a Signals course was 69.40 percent (n = 5,134 students); students in the same cohort with at least one Signals course had a retention rate of 87.42 percent (n = 1,518 students). Even more encouragingly, students with two or more Signals courses had a retention rate of 93.24 percent (n = 207 students).8 The retention numbers skyrocketed for Signals courses, yet the incoming student SAT scores for students in two or more Signal courses was more than 50 points lower than for the no-courses group.9 What this tells us is that students who are less prepared for college — as measured solely by standardized test score — are retained by and graduated from Purdue at higher rates than their better-prepared peers after having one or more courses in which Signals was used. In short, by receiving regular, actionable feedback on their academic performance, students were able to alter their behaviors in a way that resulted in stronger course performance, leading to enhanced academic performance over time.10 Student perceptions of Signals were equally encouraging: 89 percent of students using Signals had a positive experience, and 74 percent said "their motivation was positively affected" by it. 11 The Signals program's success is directly attributable to using big data to offer direct feedback. The use of analytics to predict academic success in Signals and other programs compels many institutions to examine the ethical issues of teaching, including accountability and the distribution of resources. With access to these predictive formulas,

		faculty members, students, and institutions must confront their responsibilities related to academic success and retention, elevating these key issues from a "general awareness" to a quantified value. The ethical issues related to academic success and retention are probably best examined as "what to do" with the new knowledge, rather than "what is the right answer" to the treatment of big data sets.12 That is, the focus is pragmatic, examining how to use information about the likelihood of academic success in a meaningful and effective manner. Likewise, the question of faculty member, student, and institutional responsibility should be examined in terms of: what the options are, where we are at the time, and what we can do. "Knowing" the likelihood of academic success forces each party to examine the issues; participation is unavoidable. However, using academic data does not trap faculty members into a particular approach; the question of how to use data has many answers and must be weighed against higher education's changing environment and the shifting expectations of key stakeholders"
MODEL VIEW CULTURE 2014: This Tweet Called My Back Cooky, Linabary, Corple 2018: *Navigating Big Data dilemmas: Feminist holistic reflexivity in social media research*	Lighed	"Big social media data, however, are controlled by companies that privilege corporate, governmental, and private research firms, which poses challenges for researchers wishing to access these data. Moreover, institutional review boards have struggled to develop policies regarding context-specific ethical practices for online research. As a result, scholars have begun to consider the ethical questions of online Big Data research, and Big Data research of social media specifically (e.g., boyd and Crawford, 2012; Brock, 2015; Leurs, 2017; Zwitter, 2014). Increasing consideration is being given to the methodological implications for knowledge production and dissemination, the implications for those who share their experiences, thoughts, and opinions (‘‘data points’’) via social media and online networks, how and whether social media constitutes a ‘‘public space,’’ and what informed consent means for social media users" (Cooky et al 2014, p. 1)

		"In December 2014, a group of activist women of color published a collective statement announcing an organized social media blackout, which was in response to the abuse, appropriation, and de-legitimation of their digital labor" (Cooky et al 2014, p. 1) "This paper draws attention to the importance of feminist ethics for Big Data social media research, as it is sensitive to issues of power, context, and subjugated knowledges, each of which we argue must be central considerations" (Cooky et al 2014, p. 9) "We are Black Women, AfroIndigenous and women of color who have organized a social media Blackout" (Collected Authors, 2014) "There is a refusal to legitimize the words of women of color without the backing of academia, established media, and non-profit monikers. How do we then legitimize the lens with which marginalized women of color view their lives and the spaces where they are actually allowed to assert their agency? Currently the tools women of color use to engage a movement that has long viewed them as silent subjects, relegated to frying chicken and frybread for the real movers, are devalued at best—and threatened after mined for content at worst. All of this and more take place to the tune of stalking, plagiarism, and an outright refusal to look at the interpersonal violence that we face as a result. Still, no one can quell our concern about how it is that we can expect to be respected and kept safe a physical movement space if you won't respect and keep us safe in a digital one" (Collected Authors, 2014: para. 8)
Taylor, N. 2016: **The ethics of big data as a public good: which public? Whose good?**	Godgørenhed	"The value of digital data on low-income areas to these sectors has given rise to calls for 'data philanthropy' [4], where corporations donate data to non-profit actors through development intermediaries, such as the UN. There are some high-profile examples of data philanthropy, for instance, where Twitter has granted access to its data to UN Global Pulse [5]. However, despite

UN Global Pulse: **Twitter and UN Global Pulse Announce Data Partnership**		their data's value to the development sector, mobile operators are proving reluctant to share call detail records (CDRs) under the kind of general principles that are being discussed by development and humanitarian actors. This paper will ask whether the argument for 'data as a public good' fits with the reality of big data, using two case studies where mobile operators have shared CDRs for development and humanitarian purposes [...] To explore the different considerations in play with data sharing, this paper will compare the experiences of two firms that have shared mobile phone CDRs for research purposes: Telenor (based in Norway) and Orange (based in France)." (p. 2)
Zimmer 2018: 'But the data is already public'': on the ethics of research in Facebook	Godgørenhed; Privatliv	"In 2008, a group of researchers publicly released profile data collected from the Facebook accounts of an entire cohort of college students from a US university. While good-faith attempts were made to hide the identity of the institution and protect the privacy of the data subjects, the source of the data was quickly identified, placing the privacy of the students at risk" (p. 1) (("Using this incident as a case study, this paper articulates a set of ethical concerns that must be addressed before embarking on future research in social networking sites, including the nature of consent, properly identifying and respecting expectations of privacy on social network sites, strategies for data anonymization prior to public release, and the relative expertise of institutional review boards when confronted with research projects based on data gleaned from social media"))
Stahl & Wright 2018: *Proactive Engagement with Ethics and Privacy in AI and Big Data*	Lighed	"Another example refers to the league tables of universities such as the one created by US News & World Report. While the league tables might have been created as a way of boosting the magazine's circulation, the algorithm that powered the weighting of those universities did not take into account some factors that it arguably should have done (e.g., which universities had the lowest tuition fees) and took into account other factors (which had the

		best sports facilities) that it could have ignored. It has been argued that the algorithm had unwanted consequences such as increasing the cost of university education in the US and the associated student debt. The result is an education system that favours the privileged [16, p. 65] . In other words, the US education system contributes to inequality – and all because of an algorithm that could have been designed better" (p. 6)
Jyllandsposten 2019: Data er blevet en central del af politiets efterforskning. Det stiller store krav til danske virksomheder Jyllandsposten 2019: **Politiets nye cybervåben, Pol-Intel, vækker jubel og frygt** DANSK IT's arbejdsgruppe for dataetik 2018: Dataetik : 18 dataetiske anbefalinger fra DANSK IT	Godgørenhed; Privatliv	"Data er blevet en central del af politiets efterforskning [...] Nye værktøjer har gjort det muligt for politiet at finde frem til mistænktes netværk og vennekreds gennem data. Eksperter peger dog på, at man herved involverer langt flere end tidligere og tager for givet, at dataene altid er korrekte [...] i de senere år har dansk politi selv pustet liv i gløderne med anskaffelsen af flere værktøjer - heriblandt Pol-Cam, der er et register over overvågningskameraer, og efterforskningsværktøjet Pol-Intel [...] Det, der før krævede en lang række dataudtræk i forskellige registre, kan nu klares på en enkelt platform. Udover at effektivisere politiets arbejde øger Pol-Intel mulighederne for at finde sammenhænge. Det kan eksempelvis bruges til at lave analyser af mistænktes netværk og vennekreds, da der både er adgang til politiets interne databaser og sociale medier. Systemet har dog fået kritik for at inddrage flere mennesker i politiarbejdet end tidligere. Eksperter peger desuden på, at den analysebaserede efterforskning vil stille øgede krav til de virksomheder, der indirekte eller direkte kommer til at levere data til politiet [...] Spørger man Søren Enevoldsen, er de nye værktøjer imidlertid højest nødvendige. Han er vicepolitiinspektør i Nationalt Efterforskningscenter under Rigspolitiet og peger på, at Pol-Intel gør det muligt for politiet at følge med udviklingen. »Det generelle antal af anmeldelser falder, men kompleksiteten stiger til gengæld. Derfor er vi nødt til at have nogle nye metoder, så vi kan analysere os frem og opklare kriminaliteten«, siger Søren Enevoldsen. Han bliver bakket op af centerchefen for Rigspolitiets Center for Databeskyttelse, Christian Wiese Svanberg. »It-kriminalitet, bedrageri på nettet, CEOfraud og børneporno - det er altsammen

		noget, der nu forekommer i den digitale verden og ikke i den fysiske. Det kræver, at vi har den tekniske kapacitet, kompetencerne og de retlige rammer til at løse opgaven,« siger centerchefen. Politiet må som udgangspunkt benytte Pol-Intel, når det er nødvendigt for at opklare en kriminel handling. [...] Det får Jesper Lund til at efterlyse en ny kontrolforanstaltning for politiets brug af databaseopslag. Han mener, at der er brug for at sætte de samme rammer for værktøjet, som man har for de traditionelle tvangsindgreb.»I mange tilfælde skal der være en uafhængig vurdering i form af en dommerkendelse, inden politiet kan lave et tvangsindgreb. Det betyder givetvis, at mange ting bliver filtreret fra, enten fordi kravet til kriminalitetens alvor ikke er opfyldt, eller at mistankegrundlaget ikke er solidt nok. Den type uafhængig kontrol, før man går i gang, har man slet ikke i Pol-Intel,« siger han" (Jyllands-Posten december 2019) "Politiets nye cybervåben, Pol-Intel, vækker jubel og frygt. Mens politiet jubler over sit nye cybervåben, frygter flere kritikere for lovlydige borgeres privatliv" (Jyllands-Posten august 2018) "Data kan anvendes i rigtigt mange sammenhænge. Med den offentlige sektors øgede digitalisering kan data eksempelvis bruges som forudsigende data eller til en helt automatiseret sagsbehandling. Et eksempel på sådanne systemer er eksempelvis forudsigende politiarbejde - det vil sige politiindsatser baseret på kunstig intelligens-forudsigelser af, om en person vil begå kriminalitet. I udlandet ses en stigende tendens, hvor disse systemer tages i brug af blandt andet politiet, og også i Danmark er det for nyligt blevet aktuelt, da Rigspolitiet i foråret 2018 introducerede brugen af systemet POL-INTEL, hvor computeralgoritmer, kunstig intelligens og data fra blandt andet sociale medier bringes i spil i efterforskningsarbejdet. POL-INTEL skulle oprindeligt også have en forudsigende funktionalitet, men

		Rigspolitiet har valgt ikke at benytte denne funktion på grund af den lave kriminalitetsrate, der er i Danmark. Det interessante her er dog, at et system som dette igangsættes uden større folkelig debat eller stillingtagen til de retssikkerhedsmæssige konsekvenser. Det er afgørende, at der værnes om borgernes retssikkerhed i den digitale verden. Når mekanismer, der er af særlig indgribende karakter for den enkelte borger, indføres uden for meget stillingtagen til særligt de retssikkerhedsmæssige konsekvenser for den enkelte borger, affødes deraf tilsvarende en trussel mod gennemsigtigheden, retssikkerheden og ikke mindst tilliden til samfundet. For slet ikke at nævne truslen mod borgerens fundamentale frihedsrettigheder. Tillid går begge veje, hvorfor der ved indførelse af sådanne mekanismer tilsvarende bør indføres demokratiske processer og effektiv demokratisk kontrol. Det er også værd at bemærke, at der er en vis fare ved at lægge for meget vægt på en algoritme og sætte sin sunde fornuft i baggrunden. Det bør ske med den største forsigtighed grundet algoritmers/datas kvalitet, herunder at systemer/data kan være fejlbehæftede, og såfremt de er det, hvem opdager det så og hvordan? Det fremgår af GDPR, at Data processing should be designed to serve mankind1 og ikke omvendt. Mistillid til systemet vil kunne give borgeren incitament til mindre deling af (korrekte og ikke mindst følsomme) data med offentlige myndigheder. Forudsigende data og til dels også automatiserede afgørelser giver anledning til etiske og ikke mindst juridiske problemstillinger, som skal overvejes grundigt, inden de bliver en del af borgerens hverdag og vores samfund. Er det først sat i værk uden en grundig stillingtagen til disse problemstillinger, kan det have konsekvenser for tilliden og ikke mindst retssikkerheden for den enkelte borger" (DANSK IT 2018)
DANSK IT's arbejdsgruppe for dataetik 2018: Dataetik : 18 dataetiske	Sikkerhed	"Et eksempel fra den virkelige verden kan illustrere, at data ikke altid er, hvad de giver sig ud for at være: I en dansk kommune er tryghedsbesøg inden for ældreplejen et kendt og udbredt fænomen. Et tryghedsbesøg er et besøg, hjemmeplejen aflægger hos en medborger, som egentlig ikke har behov for nogen særlig

anbefalinger fra DANSK IT		eller praktisk bistand på dét tidspunkt, men som klart trives bedst med ikke at føle sig ensom og forladt. Der kan for kommunen være en udmærket business case i sådan et tryghedsbesøg, eftersom den pågældende borger til gengæld eksempelvis undlader at bruge sit nødkald, typisk uden for normal arbejdstid, eller på anden måde medfører ekstra opgaver for hjemmeplejen. Imidlertid findes ydelsen "tryghedsbesøg" desværre ikke i kommunens ydelseskatalog. Det kan med andre ord ikke lade sig gøre at bevilge et tryghedsbesøg. Personalet skal dog stadig vælge en ydelse, når de registrerer deres besøg. De kalder det derfor bare noget andet. Forestiller man sig nu, at nogle kollegaer, fx i økonomiafdelingen, lidt for ukritisk baserer sig på de data, der kommer ud af ovenstående lille udpluk, siger det sig selv, at resultaterne meget vel kan være grundigt misvisende – for under hvilke kategorier er tryghedsbesøget nu blevet registreret?"
DANSK IT's arbejdsgruppe for dataetik 2018: Dataetik : 18 dataetiske anbefalinger fra DANSK IT	Godgørenhed; Privatliv	"Et eksempel på et dilemma: Staten ønsker at forebygge og behandle lidelser, der er belastende både menneskeligt og økonomisk - såsom psykiske lidelser. Borgerne har en lang række sundhedsdata liggende på borger.dk. Der er samtidigt voksende og indbringende marked for at udvikle medicin og behandlingsmetoder for stat såvel som private virksomheder. Staten ønsker at videregive uidentificerbare oplysninger til både private og offentlige institutioner, der har et formål inden for forebyggelse og behandling. Oplysningerne er både på levende og afdøde, herunder oplysninger vedr. dødsårsager såsom selvmord. Arvelighed bliver et fokusområde: Der er måske genetiske spor, der - hvis de bliver identificeres tidligt - kan bidrage til en effektiv forebyggelse af menneskelige lidelser og tunge behandlingsbudgetter. Samkøring af registre er muligt, hvis man må benytte identificerbare data. Hvem skal give tilladelse til anvendelse af identificerbare data? Skal en familie med et tragisk selvmord som livsbagage, man har brugt år på at bearbejde, tvinges til at tage stilling til adgang til en afdød fars sundhedsdata til gavn for forskning og udvikling? Og kan familien være sikker på, at data,

		som nu kan henføres til dem selv som personer, ikke bliver brugt til andre formål, eksempelvis i forbindelse med forsikringer eller lignende?"
Forsikring & Pension 2019: Databrug og dataetik : Dilemmaer og mulige positioner for forsikrings- og pensionsbranchen	Lighed	"Case 1: Jon Cooper om Life.io: Fordele og ulemper ved personaliserede services "We focus on engaging people – building a relationship with the customer. The bi-product of that is very rich data. So first and foremost, our mission is to help individuals get more out of their insurance product. So through our platform, the carrier can engage the policy holder, set goals around their health, finances, life events and track their progress, give rewards and so on." – Jon Cooper, Co-founder & CEO, Life.io. Forretningsmodellen bygger på data fra interaktioner med den online platform samt sundhedsdata, finansielle data, livsstil, psykografiske data som værdier og interesser samt livsbegivenheder. Data indsamles fra kilder såsom Fitbitt, Iwatch, bankkonti, lægejournaler mv. samt teknologi og kunstig intelligens i form af algoritmer, der identificerer kunders sandsynlighed for nyt køb eller merkøb. Maskinanalysen er baseret på den store mængde data fra samtlige kunder, der minder om hinanden. Life.io er således et eksempel på, hvordan marketing-modellerne fra tech-giganterne (fx Amazon) bliver kopieret i forsikrings- og pensionsbranchen. Forretningsmodellen er i høj grad baseret på kundeforholdet og kundens vilje til at dele dette med platformen. Og når der er tale om følsomme data, bliver det en opgave at forklare kunden, hvorfor han eller hun bør dele det, og hvilken værdi det giver den enkelte at dele data. Life-casen illustrerer, at der er store muligheder og udfordringer knyttet til anvendelse af mere data til det dataetiske grundtema "personalisering" af risikoberegning og prisfastsættelse. Øget personalisering kan føre til mere korrekt og fair risikovurdering og prisfastsættelse, men også identificering/isolering af de svageste grupper, som er et centralt

		dataetisk dilemma knyttet til personalisering. Herhjemme er den reelle besparelse ved at dele flere data vel at mærke minimal. Det er primært ift. gamle huse og unge bilister, at der i dag vil være en mærkbar besparelse. • Casen illustrerer, hvordan der særligt på personsiden opstår etiske dilemmaer ved mere dataanvendelse ifm. risikovurdering. Men i Danmark er denne problematik ift. især sundhedsforsikring ikke lige så væsentlig som i andre lande, da vi har valgt at alle skal betale samme pris for sundhedsbehandling uanset individuelle risici. • Casen illustrerer også, hvordan det traditionelt opfattede solidaritetsprincip forstået som systematisk omfordeling ikke kan forenes med personalisering, som knytter sig til etikken omkring fairness for den enkelte og forsikringsfællesskabet. Færre data til risikoberegning vil skabe større uretfærdighed for ham, der prissættes for højt eller for de andre, hvis han prissættes for lavt. • Casen illustrerer dermed også værdiskabelsen i at anvende data til at bryde stigmatiseringer og uretmæssige segmenteringer af grupper baseret på få datapunkter. • Casen illustrerer, at høj grad af personalisering og solidaritet, forstået som usystematisk omfordeling fra ikke-skadelidte til den uheldige, kan fungere sideløbende og uden modsætninger. • Casen illustrerer de etiske problemstillinger, der rejser sig ift. temaet ”incitamenter” til dataudveksling til gengæld for fx lavere pris/besparelser. Her er der risiko for, at sådanne transaktioner vil ramme socialt skævt, men at det ikke nødvendigvis vil øge uligheden, men nærmere bekræfte en eksisterende ulighed, der allerede er i samfundet. Derudover vil økonomiske incitamenter gøre, at yderligere datadeling vil ske på områder, hvor gevinsten af navnlig dynamiske sensordata vil være mærkbar. • Casen peger ligeledes på, hvordan en sådan tilskyndelse til datadeling kan ændre puljerne inden for det eksisterende forsikringsmarked, hvilket vil betyde, at der navnlig inden for

		person og adfærdsdata kan opstå en differentiering fra høj- til lavrisikogrupper, hvor flere små puljer vil betale mindre, mens færre store højrisikogrupper vil betale mere. Den differentiering vil ske på personområdet, men sandsynligvis i mindre grad end det ses på tingsskade. De relevante etiske diskussioner vil være begrænset til personområdet, hvor det vil dreje sig om problematikker vedr. A- og B-hold. • Casen illustrerer i samme omgang, at transparens og personligt valg spiller en central rolle for udviklingen af markedet samt konsekvenserne for de svageste og mest risikable grupper. Egenkontrol over data bliver i den forbindelse et vigtigt dataetisk tema, og en forudsætning for en potentiel markedsudvikling mod mere og mere præcist beregnede grupper i bunden af risikoskalaen og større grupper i toppen. Det er også sådan markedsudviklingen har været i Danmark over de sidste mange år. • Casen giver også anledning til refleksion over problemet med ”frivillighedens illusion”, hvor kunden enten kan dele data med samtykke eller modtage et meget ringe alternativ. Her er der ikke tale om reelt frit valg, hvilket er et centralt element, hvis branchen etisk forsvarligt skal kunne anvende flere data til bedre risikoberegning. • Casen åbner for muligheden for, at forsikringsprodukter kan blive tilgængelige for langt flere grupper, end det reelt er i dag, hvis forsikringstagere motiveres til at forbedre deres risici og blive ’forsikringsbare’. Den store potentielle dataetiske udfordring opstår imidlertid i samme øvelse, eftersom der kan opstå mulighed for ikke-forsikringsbare personer og grupper (fx ”røde huse”). Grupper, vi som samfund kan have en etisk forpligtigelse til at tilbyde en løsning. Det kunne fx være et max-bånd alle skal forsikres indenfor i et land/EU, eller at vi fælles over skatten finansierer opkøb af udsatte boliger og tilbyder genhusning til husenes beboere"

Forsikring & Pension 2019: Databrug og dataetik : Dilemmaer og mulige positioner for forsikrings- og pensionsbranchen	Etik	"Case 2: Fordele og ulemper ved at bruge data til at bekæmpe svindel [...] Shift Technology er en InsurTech-virksomhed med hovedkontor i Paris. Virksomheden blev grundlagt i 2014 og beskæftiger i alt 90 medarbejdere. Shift Technology tilbyder forsikringsselskaber avancerede anti-svindel systemer. Shift har et stort antal 'data scientists' ansat, som lærer maskiner at opdage mønstre, der indikerer svindel. Disse data scientists sidder ude hos forsikringsselskaberne og taler med skade- og svindeleksperter, og lærer maskinerne den viden, de får fra eksperterne" • Shift-casen illustrerer dilemmaet mellem nytteværdi ved anvendelse af flere data til at identificere mulige svindelsager vs. unødvendig og uberettiget kontrol af de mange retskafne kunder for at fange de få svindlere. • Casen illustrerer også dilemmaet mellem forbrugerhensyn ift. mængden af data, man indsamler og registrerer sat op imod, hvor meget svindel man er villig til at acceptere, og som den retskafne kunde vel at mærke skal være med til at betale for. • Flere data skaber større nytteværdi ift. mere præcis identificering af svindlere og færre falsk positive. Men det er ikke absolut nødvendigt at have mange datakilder for at identificere mulige svindelsager. Så hvordan skal man finde den rigtige dataetiske skæring? • Casen illustrerer, at flere datapunkter og forskellige datakilder med fordel kan samkøres for at opnå endnu bedre opdagelse af svindel, uden at retskafne kunder generes unødigt. Flere data og kunstig intelligens kan altså potentielt løse konflikten mellem at varetage hensynet til kundens pengepung eller privatlivsfred, og at gøre begge dele vha. af intelligent mønstergenkendelse så der kun indsamles flere data, når der er god grund til mistanke. • Casen illustrerer, at nye dataanvendelsesmuligheder navnlig ift. identifikation af potentielle svindlere vil få stor betydning for

		skadesagsbehandling og svindel, som i et vist omfang vil smelte sammen og køre fuldautomatisk. • Casen illustrerer, at der med stigende digitalisering af skadeanmeldelsen vil være et stigende behov for at kunne imødegå svindel med systemer, samt præventivt med incitamenter til etisk stillingtagen hos kunden ifm. skadeanmeldelsen. • Slutvist illustrerer casen, at incitament til at give falske eller sande data er en væsentlig dataetisk problematik ift. skadesagsbehandlingen"
Forsikring & Pension 2019: Databrug og dataetik : Dilemmaer og mulige positioner for forsikrings- og pensionsbranchen	Etik	"Case 3, Luca Schnettler om HealthyHealth: fordele og ulemper ved forebyggelse [...] HealthyHealth er et InsurTech-firma, der blev grundlagt i 2017 af Luca Schnettler. Det er placeret i London og beskæftiger 15 medarbejdere. Formålet med virksomheden har fra start været at innovere forsikringssektoren vha. digitale redskaber, gøre forsikringskunder sundere og forhindre sundhedsskadelig adfærd. HealthyHealth bruger digitale data, social media data, medicinsk data, emotionelle data mv. fra gamification, musikpræferencer, apps og registerdata mv. til at identificere risici og risikoprofiler, og forebygge risici ved at udstede individualiserede sygdomsforebyggelsesprogrammer/sundhedsprogrammer, så fx langtidssygemeldingen ikke indtræffer" • HealtyHealth-Casen illustrerer først om fremmest, hvordan datadeling og overvågning tænder op under det ”adfærdsregulering”, som et dataetisk grundtema, der rummer både positive muligheder og udfordringer. • Casen illustrerer de mange etisk forsvarlige gevinster, ved at forsikrings- og pensionsbranchen yder præventiv rådgivning. Mange danske selskaber gør også dette i stort omfang allerede i dag.

		• Individet kan – vha. data – ændre sin adfærd og opnå lavere forsikringspræmier, bedre livsudsigter og blive en mindre belastning for fællesskabet. Det sidste repræsenterer en enorm etisk udfordring ift. værdiskabende og nytte/gavn for individet og samfundet. Her spiller individets ”egenkontrol over data” ind som et væsentligt dataetisk grundtema og forudsætning for, at adfærdsregulering på baggrund af data – som individet på transparent grundlag kan vælge at delle eller lade være - er pligtetisk forsvarligt. • Adfærdsregulering kan imidlertid også være urimelig og ufornuftig evt. spildt pga. kundens manglende viden om, hvad der virker og ikke virker (kost, motion, kørsel mv.). • Casen illustrerer imidlertid muligheden af, at vi med mere og mere data kan opnå bedre og mere fornuftige individualiserede handleplaner, så kunden kan ændre adfærd på et mere korrekt grundlag. • Casen illustrerer ydermere, at ”datasikkerhed” er et dataetisk grundtema, og som kommer til at blive en afgørende faktor for mere datadeling, som sammen med IoT og flere og flere datakilder vil være afgørende for adgang til data og mere skræddersyede forsikringsprodukter samt adfærdsregulerende produkter. • Selskaberne vil i det scenarie skulle fokusere på de faktorer (transparens, fordele, sikkerhed mv.), der påvirker datadeling, og ikke mindst at data er sand. • For at kunder vil acceptere at dele data med selskaber i det omfang, som casen illustrerer er muligt, vil en værdikædeudvidelse ud i navnlig sundhedssektoren være en fordelagtig mulighed for branchen"

Stahl & Wright 2018: Proactive Engagement with Ethics and Privacy in AI and Big Data	Etik	"Ethical issues often are complex with unintended consequences. For example, predictive policing algorithms focus on “hot spots” or, if you like, street crime. They identify where crimes have occurred and at what times, so that police forces can deploy officers to those areas at those times in an effort to prevent crimes before they occur. The use of such algorithms raise several ethical issues. One is that they reinforce an existing spiral – i.e., the police will arrest more people from those hot spot neighbourhoods because more police have been deployed in those areas rather than other areas. A broader concern is that such approaches to policing can change the overall use of law enforcement resources in ways which may be undesirable or suboptimal" (p. 6)
Weekendavisen 2019: Uansvarlige algoritmer	Autonomi; etik; værdighed	"Hvorfor spilde en uddannelse eller efteruddannelse på en person, der med høj sandsynlighed ikke gennemfører – ifølge algoritmens forudsigelser? [...] Data om adfærd virker fristende objektive til at træffe afgørelser ud fra. Det sparer både besvær og udgifter, tror man. Og det kan være rigtigt, men er i sig selv også datadeterminisme, som Silicon Valley-profeter fremturer med. Her er menneskelig refleksion og dømmekraft overladt til de langt klogere maskiner. »The end of theory«, som et slogan lyder, hvormed man mener, at med nok data kan man forudsige alt helt præcist. Det er både dumt, naivt og ikke mindst etisk problematisk" (Weekendavisen 2019)
Weekendavisen 2019: Uansvarlige algoritmer	Autonomi; etik; værdighed	"Hvorfor skal jeg ansætte en medarbejder, hvis min dataanalyse kan afsløre, at hun nok snart melder sig syg? [...] Data om adfærd virker fristende objektive til at træffe afgørelser ud fra. Det sparer både besvær og udgifter, tror man. Og det kan være rigtigt, men er i sig selv også datadeterminisme, som Silicon Valley-profeter fremturer med. Her er menneskelig refleksion og dømmekraft overladt til de langt klogere maskiner. »The end of theory«, som et slogan lyder, hvormed man mener, at med nok data kan man

		forudsige alt helt præcist. Det er både dumt, naivt og ikke mindst etisk problematisk" (Weekendavisen 2019)
Weekendavisen 2019: Uansvarlige algoritmer	Værdighed	"Hvorfor igangsætte en dyr behandling af en patient, når sandsynligheden for, at hun snart dør, er meget stor – som big dataanalysen viser? [...] Data om adfærd virker fristende objektive til at træffe afgørelser ud fra. Det sparer både besvær og udgifter, tror man. Og det kan være rigtigt, men er i sig selv også datadeterminisme, som Silicon Valley-profeter fremturer med. Her er menneskelig refleksion og dømmekraft overladt til de langt klogere maskiner. »The end of theory«, som et slogan lyder, hvormed man mener, at med nok data kan man forudsige alt helt præcist. Det er både dumt, naivt og ikke mindst etisk problematisk" (Weekendavisen 2019)
Weekendavisen 2019: Uansvarlige algoritmer	Fairness; retfærdighed	"Hvorfor ikke slippe en fængslet fri, når data viser, at hans type netop ikke gentager forbrydelsen? [...] Data om adfærd virker fristende objektive til at træffe afgørelser ud fra. Det sparer både besvær og udgifter, tror man. Og det kan være rigtigt, men er i sig selv også datadeterminisme, som Silicon Valley-profeter fremturer med. Her er menneskelig refleksion og dømmekraft overladt til de langt klogere maskiner. »The end of theory«, som et slogan lyder, hvormed man mener, at med nok data kan man forudsige alt helt præcist. Det er både dumt, naivt og ikke mindst etisk problematisk" (Weekendavisen 2019)
Weekendavisen 2019: Uansvarlige algoritmer	Etik	"Hvorfor rådgive min kunde til investering i en forening, hvor jeg via datainformerede dashboards kan se, at den ikke performer? [...] Data om adfærd virker fristende objektive til at træffe afgørelser ud fra. Det sparer både besvær og udgifter, tror man. Og det kan være rigtigt, men er i sig selv også datadeterminisme, som Silicon Valley-profeter fremturer med. Her er menneskelig refleksion og dømmekraft overladt til de langt klogere maskiner. »The end of theory«, som et slogan lyder, hvormed man mener, at med nok

		data kan man forudsige alt helt præcist. Det er både dumt, naivt og ikke mindst etisk problematisk" (Weekendavisen 2019)
Dagbladet Roskilde 2019: Dine data skal være dine - og dine alene	Fairness	"Hun fortæller, at både hoteller, flyselskaber og biludlejning bruger ' differentieret prissætning'. Men det begynder også at blive brugt af nogle forsikringsselskaber i USA, såkaldt mikrotarifering. -Nogle amerikanske forsikringsselskaber er begyndt at sætte præmier og selvrisiko afhængig af din adfærd. Det er måske fair nok, når det for eksempel gælder bilforsikringer, hvor du kan risikerer at få en højere selvrisiko, hvis du har kørt for stærkt, og trackeren i bilen har registreret det. Men hvad med sådan noget som en sundhedsforsikring, siger hun og påpeger, at visse selskaber tracker, altså følger med i, hvordan folks helbred er, for at vurdere om de må få en billigere sundhedsforsikring. Hvilket kan være fint, hvis du er under 40 år og i god form, men når du bliver ældre, og kroppen naturligt begynder at svækkes, så kan det være svært at få råd til en forsikring, forklarer Pernille Tranberg og spørger: -Synes du, det er fedt?" (Sjællandske Slagelse 2020)
Propublica 2017: California to Investigate Racial Discrimination in Auto Insurance Premiums	Fairness	"California to Investigate Racial Discrimination in Auto Insurance Premiums. The state's insurance department is following up on our findings that eight auto insurers charge more in minority neighborhoods than in other neighborhoods with similar risk"
Dagbladet Roskilde 2019: Dine data skal være dine - og dine alene	Demokrati	"I Europa ønsker vi ikke et datadiktatur. Men det, som de har i USA, og som Pernille Tranberg kalder »datamonopol«, ønsker vi det?, spørger hun [...] kontrollen skal ikke ligge hos Facebook eller Google eller staten. Det skal ligge hos dig. Ellers opnår vi ikke datademokrati, fastslår Pernille Tranberg [...] Men hvordan kan vi i første omgang undgå at give vores data til de store techgiganter? - Du kan starte med at kigge dig om efter nogle alternativer. Der er ikke et stort alternativ til Facebook. Men der er masser af alternativer til Facebook Messenger, siger hun og foreslår apps som

		Signal og Wire, der er oplagte fremfor Messenger og WhatsApp. Derudover bør vi holde os fra at bruge Googles internetbrowser Chrome - som ellers op mod 90 procent af internetbrugerne benytter - og gå over til sådan nogle som Firefox eller Safari eller Cliqz. Firefox er ejet af Mozilla, en non profit-virksomhed, der arbejder for privacy og individuel selvbestemmelse, så her bliver dine data ikke gemt og delt med alle mulige andre og brugt til at profilere dig. - Firefox er »private by default« i den nyeste version. Det vil sige, at de blokerer for tracking af dig, uden du skal gøre noget. Det er skideflot, og det er dataetik, fordi de går skridtet videre end det, GDPR siger, fortæller Pernille Tranberg. Problemet er, at folk fravælger Firefox, der taber markedsandele, og i større omfang vælger Chrome, hvilket datarådgiveren ikke forstår. - Hvorfor f…. bruger vi den, det kunne jeg aldrig finde på, når der er gode alternativer, som ikke tracker dig, udbryder hun" (Sjællandske Slagelse 2020)
Dagbladet Roskilde 2019: Dine data skal være dine - og dine alene	Lighed	"[...] bruger det amerikanske retsvæsen som et andet eksempel. - Hvis data ikke er renset for bias (forudindtagede holdninger), så risikerer man det, man har gjort i årevis i USA, at man har smidt mange flere sorte end hvide i fængsel. Og det vil sige, at hvis du er sort, er chancen for at blive prøveløsladt måske meget mindre, end hvis du er hvid. Så vi gentager nogle racistiske mønstre, hvis vi ikke er forsigtige, forklarer hun" (Sjællandske Slagelse 2020)
Politiken 2020: Debat: Vi har skabt en folkeskole drevet af blinde data	Lighed; Autonomi; Sikkerhed	"DET SÅKALDTE learning analytics, eller datainformeret skoleudvikling, er det nye sort i folkeskolen, som Business Intelligence længe har været det i erhvervslivet. Men hvor erhvervslivet hurtigt reagerer, når dårlige data er i spil, fordi det truer indtjeningsgrundlaget, så er det tvivlsomt, om folkeskolen er klædt på til at håndtere data optimalt. En dataanalyse kan være lige så fejlbehæftet som analysen, der kobler frøer og hjernekapacitet [...] Og hvordan måler man den enkelte elevs progression? Det gør man med profileringsværktøjer, der gradvist vinder indpas i

		folkeskolen. Her forudsiger man elevens fremtidige læringsadfærd på baggrund af elevens nuværende standpunkt sammenlignet med historiske data over tilsvarende elevgrupper. Med den tilgang bliver det sværere at falde uden for normen, sværere at være en grim ælling, sværere at være lidt skæv engang imellem, sværere ikke at passe ind i normaliseringsmagtens forventningsskabelon. Ovenikøbet bliver det også sværere at forsvare sig mod dataanalytiske resultater og vurderinger. Det er notorisk umuligt at forsvare sig mod forudsigelser, fordi de ikke er faktuelle, men matematiske spådomme bestående af sandsynlighedsberegninger over elevens fremtidige udviklingskurve. Ressourcestærke forældre kan sikkert tune ind og sikre, at deres børn navigerer målrettet og trygt igennem algokratiets farvande. Men sårbare forældre rammes hårdere af de velmenende initiativer, der skal sikre alle børn den optimale skolegang. Det er bare svært at se, hvordan det skal ske, hvis man tager modet fra dem, der spiser frøer, og i stedet udstyrer dem med en risikoprofil, der indebærer, at nysgerrighed erstattes med ensrettende datadrevne læringsinitiativer" (Politiken Kultur januar 2020)
Slade, Sharon and Prinsloo, Paul (2013): Learning analytics: ethical issues and dilemmas	Lighed; fairness	"The location and interpretation of data It is now the case that “significant amounts of learner activity take place externally [to the institution]... records are distributed across a variety of different sites with different standards, owners and levels of access” (Ferguson, 2012, para. 6). This flags the difficulties associated with attempting to enforce a single set of guidelines relating to ethical use across such a range of sites, each with its own data protection standards, for instance. In addition, there are questions around the nature and interpretation of digital data as fully representative of a particular student (cohort). Correlations between different variables may be assumed when dealing with missing and incomplete data around usage of the institution’s learning management system (LMS) (Whitmer, Fernandes & Allen, 2012). Such assumptions may be influenced by the analyst’s own

		perspectives and result in subconsciously biased interpretations. The distributive nature of networks and the inability to track activity outside of an institution's internal systems also impacts the ability to get a holistic picture of students' life-worlds. Not only do we not have all the data, a lot of the data that we do have requires "extensive filtering to transform the 'data exhaust' in the LMS log file into educationally relevant information" (Whitmer et al, 2012, para. 22). There are implications, too, of ineffective and misdirected interventions resulting from faulty learning diagnoses which might result in "inefficiency, resentment, and demotivation" (Kruse & Pongsajapan, 2012, p.3). In addition, systematized modeling of behaviors, which necessarily involves making assumptions (e.g. regarding the permanency of students' learning contexts) , can determine and limit how institutions behave toward and react to their students, both as individuals and as members of a number of different cohorts. Ess, Buchanan and Markham (2012) concur that there is a need to consider the individual within what may be a very large and depersonalized data set even if that individual is not recognizable. Actions influenced by a cohort of which a student is a single part may still adversely impact on that student's options. In his recent article, Harvey (2012, para. 9) warns that "this process portends a reification of identities, with support allocated by association rather than individual need". Given the wide range of information which may be included in such models, there is a recognized danger of potential bias and oversimplification (Bienkowski et al, 2012; Campbell et al, 2007; May, 2011). In accepting the inevitability of this, should we also question the rights of the student to remain an individual and whether it is appropriate for students to have an awareness of the labels attached to them? Are there some labels which should be prohibited? As students become more aware of the implications of such labeling, the opportunity to opt out or to actively misrepresent certain characteristics to avoid labeling can diminish the validity of the remaining data set. Many institutions are employing learning analytics to nudge students toward study

		choices or to adopt support strategies which are assumed to offer greater potential for success (Parry, 2012), but what is the obligation for the student to either accept explicit guidance or to seek support which may be in conflict with their own preferences or study goals (Ferguson, 2012)? There is a risk of a “return to behaviorism as a learning theory if we confine analytics to behavioral data” (Long & Siemens, 2011, pp.36-38)" (p. 8)
Slade, Sharon and Prinsloo, Paul (2013): Learning analytics: ethical issues and dilemmas	Autonomi; Privatliv	Informed Consent, Privacy and the De-identification of Data Whilst students are increasingly aware of the growing prevalence of data mining to monitor and influence buying behavior, it is not clear that they are equally aware of the extent to which this occurs within an educational setting. Epling, Timmons and Wharrad (2003) discuss issues around the acceptability of student surveillance and debate who the real beneficiaries are. Use of data for non-educational purposes is flagged explicitly by Campbell et al (2007) referring to the use of student related data for fundraising, for example. Wel and Royakkers (2004) discuss the ethics of tracking and analyzing student data without their explicit knowledge. Interestingly, Land and Bayne (2005) discuss the broad acceptance of student surveillance and cite studies in which they record that the concept of logging educational activities seems to be quite acceptable to students. The notion of online privacy as a social norm is increasingly questioned (Arrington, 2010; Coll, Glassey & Balleys, 2011). Considering the general concern regarding surveillance and its impact on student and faculty privacy, Petersen (2012) points to the importance of the de-identification of data before the data is made available for institutional use, including the option to “retain unique identifiers for individuals in the data set, without identifying the actual identity of the individuals” (p.48).

		This latter point addresses the need to provide interventions for groups of students based on their characteristics or behaviors whilst ensuring their anonymity within the larger data set"
Slade, Sharon and Prinsloo, Paul (2013): Learning analytics: ethical issues and dilemmas	Fairness	"The Classification and Management of Data [...] Integral to contemplating the ethical challenges in learning analytics is a consideration of the impact of the tools used. Wagner and Ice (2012) explore the relevance of pattern recognition and business intelligence techniques in the evolving learning analytics landscape which provide scope for increased success by guiding stakeholders to "recognize the proverbial right place and right time" (p.34).While pattern recognition has huge potential for delivering custom-made and just-in-time support to students, there is a danger, as highlighted by Pariser (2011) that pattern recognition can result in keeping individuals prisoner to past choices. Pariser (2011) suggests that the use of personalized filters hints of "autopropaganda, indoctrinating us with our own ideas, amplifying our desire for things that are familiar", and that "knowing what kinds of appeals specific people respond to gives you the power to manipulate them on an individual basis" (p. 121). As we propose in the later section on ethical considerations, the algorithms used by institutions invariably reflect and perpetuate current biases and prejudices. The dynamic nature of student identity necessitates that we take reasonable care to allow students to act outside of imposed algorithms and models" (p. 9)
Kaslow, Patterson, Gottlieb 2011: Ethical dilemmas in psychologists	Privatliv; Godgørenhed	"Many psychologists search the Internet for both personal and professional information. Although various guidelines have been proposed for psychologists regarding therapeutic services provided over the Internet, few address the ethics and efficacy of gathering information about clients, students, or employees on the web. As

accessing Internet data: Is it justified?		quickly as guidelines are written, new technologies create new challenges. With the advent of social networking sites and numerous free and paid data search engines, unique dilemmas have arisen. The ready access of voluminous personal information raises perplexing questions for clinician psychologists, instructors, supervisors, and employers. An overarching consideration addressed in this article is whether in the course of one's professional activities it is ethically appropriate to conduct intentional Internet searches for information about patients, students, or employees. We discuss ethical dilemmas such as right to privacy, trust, confidentiality, informed consent, boundary violations, and best interest of the client, student, or employee. Next we provide a list of some extant electronic sources of information and offer case examples. The article concludes with recommendations that we hope will generate further dialogue and research on these perplexing issues and provide guidance on balancing situationally appropriate flexibility with the need for adopting wise parameters of professional behavior in regard to social networking activities and Internet "investigations."" (p. 1)
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Sikkerhed	"The first issue is that, in order for a visualization to be rendered from a data set, those data must be translated into some visual form—that is, what is called "the principled mapping of data variables to visual features such as position, size, shape, and color" (Heer et al., 2010, p. 67; cf. Börner, 2007; Chen, 2006; Ware, 2013). By exploiting the human ability to organize objects in space, mapping helps users understand data via spatial metaphors (Vande Moere, 2005). Data, however, take different forms, and this kind of translation requires a substantial conceptual leap that leaves its mark upon the resulting product. Considered from this perspective, then, "any data visualization is first and foremost a visualization of the conversion rules themselves, and only secondarily a visualization of the raw data" (Galloway, 2011, p. 88) [...] Despite technical sophistication, decisions about these visual encodings are

		not always carried out in an optimal way (Kostelnick, 2007, p. 285) [...] Moreover, the generative nature of visualizations as a type of inquiry is often hidden—that is, some of the decisions are dismissed later as nonanalytical and not worth documenting and reporting, while in fact they may have contributed to the ways in which the data will be seen and interpreted (Markham, 2013)." (p. 1532)
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Etik	"Another issue arises in the tension between accuracy of visualizations and their aesthetic appeal. As a matter of general practice, the rules employed to create visualizations are often weighted towards what yields a more visually pleasing result rather than directly mapping the data"
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Fairness	"[...] This provides a clear competitive advantage to large players such as Google that can rely on freely available software and yet afford significant investments in purchasing and customizing hardware and software components (Fox, 2010; Metz, 2009). This would also reintroduce the invisible divide between “haves”—those who can benefit from the ideological drive toward “more” (more data, more storage, more speed, more capacity, more results, and, ultimately, more profit or reward)— and “have nots” who lag behind or have to catch up by utilizing other means" (p. 1534)
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Ansvarlighed	"The Role of Humans: Automation or Heteromation? [...] the technologies behind Big Data projects have another aspect that differentiates them from earlier technologies of automation, namely, the changing role and scale of human involvement. While earlier technologies may have been designed with the explicit purpose of replacing human labor or minimizing human involvement in sociotechnical systems, Big Data technologies

		heavily rely on human labor and expertise in order to function [...] The shared tenet among these various projects is their critical reliance on technologies that require active human involvement on large scales. Rather than describing the use of machines in Big Data in terms of automation, perhaps we should acknowledge the continuing creative role of humans in knowledge infrastructure and call it “heteromation” (Ekbia & Nardi, 2014). Heteromation and the rise of social machines also highlights the distinction between data generated by people versus data generated about people. This kind of “participatory personal data” (Shilton, 2012) describes a new set of practices where individuals contribute to data collection by using mediating technologies, ranging from web entry forms to GPS trackers and sensors to more sophisticated apps and games. These practices shift the dynamics of power in the acts of measurement, groupings, and classifications between the measuring and the measured subjects (Nafus & Sherman, under review). Humans who have always been the measured subjects receive an opportunity to “own” their data, challenge the metrics, or even expose values and biases embedded in automated classifications and divisions (Dwork & Mulligan, 2013). Whether or not this kind of “soft resistance” will tilt the balance of power in favor of small players is yet to be seen, however" (p. 1535)
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Privatliv	"The risk of free riding is a collective action problem well known to intellectual property theorists. Resources that are costly to produce and subject to cheap duplication tend to be underproduced because potential producers have little to gain and everything to lose. The function of intellectual property laws is to create an incentive for people to produce such resources by entitling them to enjoin copyists for a limited period of time. Conventional data, however, do not meet the eligibility requirements for patent protection, and are often barred from copyright protection because commercially published data are often factual in nature (Patent Act, Copyright Act). (When conventional data have met the eligibility requirements of copyright, protection has generally been

		thin.) This facet of American intellectual property law has led to efforts by database companies to urge Congress to enact new laws that would provide sui generis intellectual-property protection to databases (Reichman & Samuelson, 1997). Big Data may mark a new chapter in the story of intellectual property, expanding it to the broader issue of data ownership. Ironically, the very methods and practices that make Big Data useful may also infuse it with subjective human judgments. As discussed earlier, a researcher's subjective judgments can become deeply infused into a data set through sampling, data cleaning, and creative manipulations such as data masking. As a result, Big Data compilations may actually be more likely to satisfy the prerequisites for copyrightability than canonical factual compilations (Mattioli, forthcoming). If so, sui generis database protection may be unnecessary. Existing intellectual property laws may also need to be adapted in order to accommodate Big Data practices. In the United Kingdom, for example, lawmakers have approved legislation that provides a copyright exemption for data mining, allowing search engines to copy books and films in order to make them searchable. U.S. lawmakers have yet to seriously consider such a change to the Copyright Act. To recap, the ethical and legal challenges brought about by Big Data present deeper issues that suggest significant changes to dominant legal frameworks and practices. In addition to privacy and data ownership, Big Data challenges the conventional wisdom of collective action phenomena such as free riding—a topic discussed in the following section.

Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Etik	Data as Asset: Contribution or Exploitation? [...] The observation of these trends reinforces the World Economic Forum's recognition of data as a new "asset class" (2013) and the notion of data as the "new oil." The caveat is that data, unlike oil, are not a natural resource, which means that their economic value cannot derive from what economists call "rent." What is the source of the value, then? This question is at the center of an ongoing debate that involves commentators from a broad spectrum of social and political perspectives. Despite their differences, the views of many of these commentators converge on a single source: "users." According to the VP for Research of the technology consulting firm Gartner, Inc., "Facebook's nearly one billion users have become the largest unpaid workforce in history" (Laney, 2012). From December 2008 to December 2013, the number of users on Facebook went from 140 million to more than one billion. During this same period of time, Facebook's revenue rose by about 1300%. Facebook's filing with the Securities and Exchange Commission in February 2012 indicated that "the increase in ads delivered was driven primarily by user growth" (Facebook, 2012, p. 50). Turning the problem of free riding on its head, these developments introduce a novel phenomenon, where instead of costly resources being underproduced because they can be cheaply duplicated (see the previous discussion of data ownership), user data generated at almost no cost are overproduced, giving rise to vast amounts of wealth concentrated in the hands of proprietors of technology platforms. The character of this phenomenon is the focus of the current debate. Fuchs (2010), coming to the debate from a Marxist perspective, has argued that users of social media sites such as Facebook are exploited in the same fashion that TV spectators are exploited.8 The source of exploitation, according to Fuchs, is the "free labor" that users put into the creation of user-generated content (Terranova, 2000). Furthermore, the fact that users are not financially compensated throws a very diverse group of people into an exploited class that Fuchs, following Hardt and Negri (2000), calls the "multitude." The nature of user contribution

		finds a different characterization by Arvidsson and Colleoni (2012), who argue that the economy has shifted toward an affective law of value “where the values of companies and their intangible assets are set not in relation to an objective measurement, like labor time, but in relation to their ability to attract and aggregate various kinds of affective investments, like intersubjective judgments of their overall value or utility in terms of mediated forms of reputation” (p. 142). This leads these authors to the conclusion that the right explanation for the explosive wealth of companies such as Facebook should be sought in financial market mechanisms such as branding and valuation. Conversely, Ekbia (forthcoming) contends that the nature of user contribution should be articulated in the digitally mediated networks that are prevalent in the current economy. The winners in this “connexionist world” (Boltanski & Ciapello, 2005) are the flexibly mobile, those who are able to move not only geographically (between places, projects, and political boundaries), but also socially (between people, communities, and organizations) and cognitively (between ideas, habits, and cultures). This group largely involves the nouveau riche of the Internet age (e.g., the founders of high-tech communications and social media companies; Forbes, 2013). The “losers” are those who have to play as stand-ins for the first group in order for the links created in these networks to remain active, productive, and useful. Interactions between these two groups are embedded in a form of organizing that can be understood as “expectant organizing”—a kind of organization that is structured with built-in holes and gaps that are intended to be bridged and filled through the activities of end users.9 Whichever of these views one considers as an explanation for the source of value of data, it is hard not to acknowledge a correlation between user participation and contribution and the simultaneous rise in wealth and poverty" (p. 1537-1538)

Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Etik; autonomi	"Data and Social Image: Compliance or Resistance? [...] In contemporary societies, the idealized self is someone who is highly independent, engaged, and self-reliant—a high-mobility person, as we saw, with a potential for re-education, reskilling, and relocation; the kind of person often sought by cutting-edge industries such as finance, medicine, media, and high technology. This "new man," according to sociologist Richard Sennett (2006), "takes pride in eschewing dependency, and reformers of the welfare state have taken that attitude as a model—everyone his or her own medical advisor and pension fund manager" (p. 101). Big Data has started to play a critical role in both propagating the image and developing the model, playing out a three-layer mechanism of social control through monitoring, mining, and manipulation. Individual behaviors, as we saw in the discussion of privacy, are under continuous monitoring through Big Data techniques. The data that are collected are then mined for various economic, political, and surveillance purposes: Corporations use the data for targeted advertising, politicians for targeted campaigns, and government agencies for targeted monitoring of all manners of social behavior (health, finance, criminal, security, etc.). What makes these practices particularly daunting and powerful is their capability in identifying patterns that are not detectable by human beings, and are indeed unavailable before they are mined (Chakrabarti, 2009). Furthermore, these same patterns are fed back to individuals through mechanisms such as recommender systems, creating a vicious cycle of regeneration that puts people in "filter bubbles" (Pariser, 2012). These properties of autonomy, opacity, and generativity of Big Data bring the game of social engineering to a whole new level, with its attendant benefits and pitfalls. This leaves the average person with an ambivalent sense of empowerment and emancipatory self-expression combined with anxiety and confusion. This is perhaps the biggest dilemma of contemporary life, which rarely disappears from the consciousness of the modern individual" (p. 1538)

Politiken 2020: Tech-ekspert: Virksomheder tjener milliarder på dine data. Systemet bør revolutioneres, så du selv kan bestemme over dine dyrebare data	Fairness	"Dine data er blevet gemt og solgt videre til en annoncør, der nu ved, at du er interesseret i varen. Dermed kommer varen frem igen og igen, mens du kedsommeligt scroller ned gennem dine venners babyopslag og madvideoer. Det er selvsagt en effektiv måde at annoncere på. Virksomheder som Facebook og Google skovler penge i kassen ved at høste data om dig og dine præferencer for at sælge dem videre til annoncørerne. Men du får ikke selv en krone for det. Hvorfor egentlig ikke? »Vores data har skabt så meget værdi, at ud af de 10 mest værdifulde virksomheder i verden tjener størstedelen penge enten direkte eller indirekte på brugernes data. Så mit synspunkt er, at hvis data skaber så meget værdi, hvorfor skal værdien så være samlet hos nogle få virksomheder?«, siger Jennifer Zhu Scott fra sit hjem i Hongkong via coronatidens foretrukne kommunikationsmiddel, Zoom" (Politiken juli 2020)
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Etik	"Saving the Phenomena: Causal Relations or Statistical Correlations? The proponents, who claim that “correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all” (Anderson, 2008, para. 19), are, in some manner, questioning the viability of the Common Cause Principle, and calling for an end to theory, method, and the “old ways.” [...] The defenders of the “old ways,” on the other hand, respond in various manners. Some, such as microbiologist Carl Woese, sound the alarm for a “lack of vision” in an engineering biology that “might still show us how to get there; [but that] just doesn’t know where ‘there’ is” (Woese, 2004, p. 173; cf. Callebaut, 2012, p. 71). Others, concerned about the ahistorical character of digital research, call for a return to the “sociological imagination” to help us understand what techniques are most meaningful, what information is lost, what data are accessed, etc. (Uprichard, 2012, p. 124). Still others suggest an historical outlook, advocating a focus on “long data”—that is, on data sets with a “massive historical sweep” (Arbesman, 2013, para. 4) [...] A more

		conciliatory view comes from those who cast doubts on “using blind correlations to let data speak” (Harkin, 2013, para. 7) or from the advocates of “scientific perspectivism,” according to whom “science cannot as a matter of principle transcend our human perspective” (Callebaut, 2012, p. 69; cf. Giere, 2006). By emphasizing the limited and biased character of all scientific representations, this view reminds us of the finiteness of our knowledge, and warns against the rationalist illusion of a God’s-eye view of the world" (p. 1529)
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Etik	"Saving the Appearance: Productions or Predictions? Prediction is the hallmark of Big Data. Big Data allows practitioners, researchers, policy analysts, and others to predict the onset of trends far earlier than was previously possible. This ranges from the spread of opinions, viruses, crimes, and riots to shifting consumer tastes, political trends, and the effectiveness of novel medications and treatments. Many such social, biological, and cultural phenomena are now the object of modeling and simulation using the techniques of statistical physics (Castellano, Fortunato, & Loreto, 2009) [...] In brief, an historical trend of which Big Data is a key component seems to have propelled a shift in terms of success criterion in science from causal explanations to predictive modeling and simulation. This shift, which is largely driven by forces that operate outside science (see the forthcoming Political Economy section), pushes the earlier trend—already underway in the 20th century, in breaking the bridge between phenomena and appearances—to its logical limit. If 19th-century science strove to “save the phenomenon” and produce the appearance from it through causal mechanisms, and 20th-century science sought to “save the appearance” and let go of causal explanations, 21st-century science, and Big Data in particular, seems to be content with the prediction of appearances alone. The interest in the prediction of appearances, and the garnering of evidence needed to support them, still stems from the rather narrow deductive-

		nomological model, which presumes a law-governed reality. In the social sciences, this model may be rejected in favor of other explanatory forms that allow for the effects of human intentionality and free will, and that require the interpretation of the meanings of events and human behavior in a context-sensitive manner (Furner, 2004). Such alternative explanatory forms find little space in Big Data methodologies, which tend to collapse the distinction between phenomena and appearances altogether, presenting us with structuralism run amok. As in the previous eras in history, this trend is going to have its proponents and detractors, spread over a broad spectrum of views. There is no need to ask which of these embodies "real" science, for each one of them bears the seeds of its own dilemmas" (p. 1530)
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Etik	"Some humanists and social scientists have expressed concern that "playing with data [might serve as a] gateway drug that leads to more-serious involvement in quantitative research" (Nunberg, 2010, para. 12), leading to a kind of technological determinism that "brings with it a potential negative impact upon qualitative forms of research, with digitization projects optimized for speed rather than quality, and many existing resources neglected in the race to digitize ever-increasing numbers of texts" (Gooding, Warwick, & Terras, 2012, para. 6–7). The proponents, on the other hand, drawing an historical parallel with "the domestication of human mind that took place with pen and paper," argue that the shift doesn't have to be dehumanizing: "Rather than a method of thinking with eyes and hand, we would have a method of thinking with eyes and screen" (Berry, 2011, p. 22) [...] Proponents of a "pure" quantitative approach herald the autonomy of the data from subjectivity: "The data speaks for itself!" However, human intervention at each step undermines the notion of a purely objective Big Data science [...] "It would be nice if all of the data which sociologists require could be enumerated," William Cameron wrote in 1963, "Because then we could run them through IBM machines and draw charts as the economists do. . . [but] not

		everything that can be counted counts, and not everything that counts can be counted" (p. 13). These words have a powerful resonance to those who observe the intermingling of raw numbers, mechanized filtering, and human judgment in the flow of Big Data" (p. 1530-1531)
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Gennemsigtighed; Privatliv	"Data making involves multiple social agents with potentially diverse interests. In addition, the processes of data generation remain opaque and under-documented (Helles & Jensen, 2013). As rich ethnographic and historical accounts of data and science in the making demonstrate, the data that emerge from such processes can be incomplete or skewed. Anything from instrument calibration and standards that guide the installation, development, and alignment of infrastructures (Edwards, 2010) to human habitual practices (Ribes & Jackson, 2013) or even intentional distortions (Bergstein, 2013) can disrupt the "rawness" of data. As Gitelman and Jackson (2013) point out, data need to be imagined and enunciated to exist as data, and such imagination happens in the particulars of disciplinary orientations, methodologies, and evolving practices. Having been variously "pre-cooked" at the stages of collection, management, and storage, Big Data does not arrive in the hands of analysts ready for analysis. Rather, in order to be "usable," it has to be "cleaned" or "conditioned" with tools such as Beautiful Soup and scripting languages such as Perl and Python (O'Reilly Media, 2011, p. 6). This essentially consists of deciding which attributes and variables to keep and which ones to ignore (Bollier, 2010, p. 13)—a process that involves mechanized human work, through services such as Mechanical Turk, but also interpretative human judgment and subjective opinions that can "spoil the data" (Andersen, 2010; cf. Bollier, 2010, p. 13). These issues become doubly critical when personal data are involved on a large scale and "de-identification" turns into a key concern—issues that are further exacerbated by the potential for "re-identification," which, in turn, undermines the thrust in Big Data research toward "data liquidity." The dilemma between the ethos

		of data sharing, liquidity, and transparency, on the one hand, and risks to privacy and anonymity through reidentification, on the other, emerges in such diverse areas as medicine, location-tagged payments, geo-locating mobile devices, and social media (Ohm, 2010; Tucker, 2013). Editors and publishing houses are increasingly aware of these tensions, particularly as organizations begin to mandate data sharing with reviewers and, following publication, with readers (Cronin, 2013)" (p. 1531)
Ekbia et. al. (2014): ADVANCES IN INFORMATION SCIENCE: Big Data, Bigger Dilemmas: A Critical Review	Privatliv	"Privacy Concerns: To Participate or Not? [...] The concept of privacy as a generalized legal "right" was introduced in an 1890 Harvard Law Review article written by Samuel Warren and Louis Brandeis—both of whom would later serve as United States Supreme Court Justices (Warren & Brandeis, 1890). "The right to be let alone," the authors argued, is an "inevitable" extension of age-old legal doctrines that discourage intrusions upon the human psyche (p. 195). These doctrines include laws forbidding slander, libel, nuisances, assaults, and the theft of trade secrets (pp. 194, 205, 212). Since the time of Warren and Brandeis, new technologies have led policymakers to continually reconsider the definition of privacy. In the 1980s and 1990s, the widespread adoption of personal computers and the Internet led to the enactment of statutes and regulations that governed the privacy of electronic communications (Bambauer, 2012, p. 232). In the 2000s, even more laws were enacted to address, inter alia, financial privacy, healthcare privacy, and children's privacy (Schwartz & Solove, 2011, p. 1831). Today, Big Data presents a new set of privacy concerns in diverse areas such as health, government, intelligence, and consumer data [...] The common theme that emerges from the foregoing examples is that vastly heterogeneous types of data can be generated, transferred, and analyzed without the knowledge of those affected. Such data are generated silently and often put to unforeseen uses after they have been collected by known and unknown others,

		implicating privacy but also leading to second-order harms, such as "profiling, tracking, discrimination, exclusion, government surveillance and loss of control" (Tene & Polonetsky, 2012, p. 63). Ironically, the protection of privacy, as well as its violation, depends on technology just as much as it depends on sound public policy. "Masking" that seeks to obfuscate personally identifying information while preserving the usefulness of underlying data, for instance, employs sophisticated encryption techniques (El Emam, 2011). Despite its sophistication, however, it has been shown to be vulnerable to re-identification, leading Ohm (2010) to opine that "[r]eidentification science disrupts the privacy policy landscape by undermining the faith that we have placed in anonymization" (p. 1704). Another technological solution to the privacy puzzle—namely, using software to meter and track the usage of individual parcels of data—requires individual citizens and consumers to tag their data with their privacy preferences. This approach, which was put forth by the World Economic Forum, portrays a future in which banks, governments, and service providers would supply consumers with personally identifiable data collected about them (World Economic Forum, 2013, p. 13). By resorting to (meta)data to protect data, though, this "solution" puts the onus of privacy protection on individual citizens. As such, it reintroduces in stark form the old dilemma of the division of labor and responsibility between the public and private spheres" (p. 1536)
Stahl & Wright 2018: *Proactive Engagement with Ethics and Privacy in AI and Big Data*	Godgørenhed; privatliv; sikkerhed	"The reason for the current prominence of SIS [eg. Google's search engine, Google Translate, Amazon's recommendation system, Amazon's Alexa home assistant, Facebook's likes, smart phones with GPS tracking and many other smart phone apps, predictive policing systems, automated share-dealing, healthcare and surgery robots, personal fitness applications, virtual and augmented reality and many others, ranging from social network data analysis for advertising to traffic data prediction for energy conservation] is that they promise solutions to many social problems and challenges. The European Commission [1] , for example,

		formulated a strategy promoting smart, sustainable and inclusive growth as a means of overcoming the financial crisis of 2008, the after-effects of which are still visible and driving European policy decisions. In addition to these ongoing effects, Europe faces various challenges, ranging from demographic change to migration, social inclusion, health care, skills and education. It is part of the accepted public narrative that many of these challenges require intelligent and specific solutions. SIS appear to be one way of achieving these policy goals. They have the potential to generate new sources of income, improve processes, provide bespoke solutions that will improve work flow and user satisfaction. At the same time, these novel SIS raise significant concerns. Privacy and data protection are the most obvious issues but they are far from the only ones. Concerns range from questions of fairness and hidden biases in big data all the way to the possibility of truly autonomous machines that may harm or kill people, but that may also be subjects of ethical rights. A key topic of debate is the social consequence that SIS may have on the future of work and employment"
Herschel & Miori 2017: *Ethics & Big Data* / Herold 2014: 10 Big Data Analytics Privacy Problems	Privatliv	"Herold [6] has identified a number of important privacy risk associated with Big Data and Big Data analytics. For example, with so much data and with powerful analytics, it may be impossible to remove the ability to identity individuals, if there are no rules established for the use of anonymized data files. For example, if one anonymized data set was combined with another completely separate data base, without first determining if any other data items should be removed prior to combining to protect anonymity, it is possible that individuals could be re-identified. The important and necessary key step she says that is usually missing is establishing the rules and policies for how anonymized data files can be combined and used together. She also notes that if data masking is not used appropriately, Big Data analysis could easily reveal the actual individuals whose data has been masked"

Herschel & Miori 2017: *Ethics & Big Data* / Herold 2014: 10 Big Data Analytics Privacy Problems	Privatliv; Autonomi	"Herold warns that Big Data can be used to try to influence and drive behaviors. What is implied here is that Big Data can be used by organizations to make a much wider variety of decisions that do not take into account the privacy of the individuals whose personal data is being exploited. This problem is further aggravated by the fact that that Big Data can be used to fill in gaps in information about individuals. This can occur because the collection of Big Data from online transactions and the Internet of Things oftentimes affords firms the opportunity to expand their knowledge about an individual without their knowledge or consent. In effect this means that decision making is oftentimes being transferred away from individual decisions that have knowable outcomes and replaced by actions derived from Big Data analytics which may have unintended consequences for many. Zoldan [7] argues that many times the Big Data being utilized for decision making is not always correct as it is oftentimes incomplete, biased and/or missing context. Despite this, organizations frequently have a false sense of confidence in the data, since there is so much Big Data available. This is especially problematic if Big Data algorithms are drawing inaccurate conclusions about customer identities and behavior based on flawed data"
Herschel & Miori 2017: *Ethics & Big Data*	Lighed	"Other potential problems associated with Big Data analysis are signal error and confirmation bias. Signal error occurs when large gaps of data have been overlooked. Confirmation bias is the phenomena that data is selectively used to confirm a preexisting viewpoint, while disregarding the data that refutes it. The point being made by Zoldan is that the use of Big Data necessarily requires skepticism and caution to avoid statistical false positives and incorrect findings that may lead to bad decisions and unintended risk for both the organizations and its customers"

Herschel & Miori 2017: *Ethics & Big Data* / Reference til Zwitter 2014: Big Data Ethics	Autonomi	"Zwitter [8] argues that Big Data has the effect of shifting the focus of ethics away from the individual's ability to make moral judgements on some notion of right or wrong as well as their accountability. Instead, Big Data requires an examination of those that have control over Big Data, because Big Data can be used to target and manipulate people to consume or behave in a certain way. Big Data stakeholders wield a significant amount of power because they control the collection and the utility of Big Data, employing data derived knowingly or unknowingly from others. Big Data can have the effect of reducing knowable outcomes of actions, while increasing unintended consequences. Therefore, Zwitter contends that Big Data fundamentally changes the nature of ethical debates by redefining what power is and where it lies and the extent to which free will in fact guides one's actions"
Herschel & Miori 2017: *Ethics & Big Data* / Richards and King 2014: Big Data Ethics	Privatliv; Gennemsigtighed	"Richards and King [9] note that large datasets are being mined for important predictions that often yield surprising insights. They assert that because of Big Data and the analytics used to examine it, all kinds of human activities and decisions are beginning to be influenced by Big Data predictions, including dating, shopping, medicine, education, voting, law enforcement, terrorism prevention, and cybersecurity. Yet while this is occurring, individuals have little idea concerning what data is being collected, let alone shared with third parties. Hence, Richards and King assert that existing privacy protections focused on managing personally identifying information are not enough when secondary uses of Big Data sets can reverse engineer past, present, and even future breaches of privacy, confidentiality, and identity. They note that Big Data efforts find many of the most revealing personal data sets such as call history, location history, social network connections, search history, purchase history, and facial recognition and much of this this information is are already in the hands of governments and corporations. And the collection of these and other data sets is only accelerating. Richards and King conclude that Big Data is producing

		increased institutional awareness and power that requires the development of Big Data ethics to protect individual rights"
Herschel & Miori 2017: *Ethics & Big Data* / Reference til Culnan & Williams 2009: How ethics can enhance organizational privacy: lessons from the choicepoint and TJX data breaches	Privatliv; Sikkerhed	"Culnan and Williams [10] have noted the potential for abuses of informational reuse and unauthorized data that could result in privacy problems. They state that information reuse involves organizations making legal decisions about new uses for the personal information they have collected, while unauthorized access involves employees viewing personal information they are not authorized to view. Both activities, information reuse and unauthorized access, can potentially threaten an individual's ability to maintain a condition of limited access to his/her personal information, harm individuals, and subsequently threaten the organization's legitimacy in its interactions with consumers, shareholders, and regulators. Privacy harms resulting from unauthorized access can include a breach of confidentiality and trust, or the financial harm to individuals from identity theft or identity fraud. Unfortunately, Big Data has only enhanced the potential for such issues"
Herschel & Miori 2017: *Ethics & Big Data* / Nunan & Di Domenico 2013: Market research and the ethics of big data	Privatliv	"When examining the implications of Big Data on market research, Nunan and Di Domenico [11] identified privacy challenges created by the use of Big Data. The first issue they documented arises from different sets of data that would not previously have been considered as having privacy implications concerns being combined in ways that then threaten privacy. They call this the unintended use paradox. One example cited is the discovery by researchers who used publicly available information and photographs from Facebook and, through application of facial recognition software, matched this information to identify previously anonymous individuals on a major dating site. In another example, anonymous 'de-identified' health information distributed between US health

		providers was found to be traceable back to individuals when modern analytical tools were applied. With Big Data comes the possibility of significantly changing the relationship that individuals have with the data collected about them. Moreover, because Big Data and data mining findings are derived using correlations among data, there is a higher likelihood for finding random connectedness on the basis of random commonalities. The result of this is that Big Data analyses may yield information that not only compromises privacy, but suggests random connections based on incidental occurrences"
Herschel & Miori 2017: *Ethics & Big Data* / Nunan & Di Domenico 2013: Market research and the ethics of big data / Mandinach, Parton, Gummer, and Anderson 2015: Ethical and appropriate data use requires data literacy	Privatliv	"The second privacy challenge Nunan and Di Domenico ascertained revolves around the fact that data is increasingly being collected autonomously, independent of human activity. The authors note that with the emergence of network-enabled sensors on everything from electricity and water supplies to airplanes, the volume of data created by these devices, and the speed with which the data must be analyzed means that data collected is automatically analyzed without any consideration for individual consent [...] Mandinach, Parton, Gummer, and Anderson [12] state that the ethical use of data involves knowing how to use data and how to protect privacy and maintain the confidentiality of data. Such knowledge includes how to remove identifying information from a data record and knowing who has access to data and when and how, and the process by which to release data or results. Unfortunately, oftentimes the processing of Big Data is automated, being processed by devices that using analytic algorithms that are insensitive to these issues. And even when humans are involved in the process, oftentimes the sheer volume of Big Data make such efforts impractical"
UK Government 2019: *Guidance :*	Ansvarlighed	"AI systems increasingly perform tasks previously done by humans. For example, AI systems can screen CVs as part of a recruitment process. However, unlike human recruiters, you cannot hold an AI

Understanding artificial intelligence ethics and safety		system directly responsible or accountable for denying applicants a job. This lack of accountability of the AI system itself creates a need for a set of actionable principles tailored to the design and use of AI systems"
Drew 2016: Data science ethics in government	Etik	"Of course, we know that people's views do not always match their behaviours. For example, nearly half of people across 20 countries say that they are willing to pay for increased levels of data privacy, but only a quarter of people, questioned in the same survey, say that they have taken free, basic steps to increase the privacy settings on their browser. This means that three-quarters of those who say that they would pay for additional privacy have not changed a simple setting on their computer"
Mishra, Neha 2020: *International Trade Law Meets Data Ethics: A Brave New World*	Godgørenhed; sikkerhed; privatliv; fairness	Data-driven technologies bring their own share of opportunities and challenges. The widespread adoption of Big Data Analytics and Artificial Intelligence ('AI') / Machine Learning ('ML') has transformed the global economy by creating new economic, social and technological opportunities, increasing market efficiencies, and reducing trade costs.1 Experts predict that these technologies can foster meaningful innovation and enable greater human well-being in different fields of life2 such as healthcare,3 disaster management,4 and delivery of public services.5 However, data-driven technologies are also increasingly being misused, including creating/propagating disinformation campaigns,6 prejudicing online privacy and security,7 and furthering algorithmic discrimination.8 1 WORLD TRADE ORGANIZATION, WORLD TRADE REPORT 4, 8, 16 (2018), https://www.wto.org/english/res_e/publications_e/world_trade_report18_e.pdf. 2 See generally OECD, Data-Driven Innovation for Growth and Well-Being, https://www.oecd.org/sti/ieconomy/data-driven-

		innovation.htm; FILIPPO A RASO ET AL, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y, ARTIFICIAL INTELLIGENCE & HUMAN RIGHTS: OPPORTUNITIES & RISKS 3 (2018), https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights. Several initiatives are underway for assessing the social benefits of data-driven technologies, especially AI, by organisations such as the UN (Global Partnership for Sustainable Development Data and Global Pulse), World Economic Forum (New Deal on Data), Data-Pop Alliance and International Telecommunications Union (AI for Global Good). 3 See generally Adam Stevenson, Better Health Through Analytics and Data-Driven Technology, THE HEALTH FOUNDATION (Oct. 29, 2019), https://www.health.org.uk/news-and-comment/blogs/better-health-throughanalytics-and-data-driven-technology; Mikael Hagstroem, Big Data Analytics for Inclusive Growth: How Technology Can Help Elevate the Human Condition, WORLD ECONOMIC FORUM (2015), https://reports.weforum.org/global-information-technology-report-2015/1-8-big-data-analytics-for-inclusivegrowth-how-technology-can-help-elevate-the-human-condition/; Hannah Kuchler, Google AI System Beats Doctors in Detection Tests for Breast Cancer, THE FINANCIAL TIMES (Jan. 2, 2020), https://www.ft.com/content/3b64fa26-28e9-11ea-9a4f-963f0ec7e134. 4 Kylie Wiggers, Google's AI Predicts Local Precipitation Patterns "Instantaneously", VENTURE BEATS (Jan. 13, 2020), https://venturebeat.com/2020/01/13/googles-ai-predicts-local-precipitation-patterns-instantaneously/ (last visited Aug. 31, 2020). 5 Lok Siying, The Impetus of Big Data for Public Service Delivery, JOURNAL OF INTERNATIONAL & PUBLIC AFFAIRS (Apr. 30, 2018), https://www.jipasg.org/posts/2019/4/30/the-impetus-of-big-data-for-public-servicedelivery.

		6 Samuel Woolley, We're Fighting Fake News AI Bots by Using More AI. That's a Mistake, MIT TECHNOLOGY REVIEW (Jan. 8, 2020), https://www.technologyreview.com/2020/01/08/130983/were-fighting-fake-news-aibots-by-using-more-ai-thats-a-mistake/. 7 DAVID LESLIE, ALAN TURING INSTITUTE, UNDERSTANDING ARTIFICIAL INTELLIGENCE ETHICS AND SAFETY 4 (2019), https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf. 8 See, eg, Joi Ito , Supposedly 'Fair' Algorithms Can Perpetuate Discrimination, THE WIRED (May 2, 2019), https://www.wired.com/story/ideas-joi-ito-insurance-algorithms/ (demonstrating the potential for discriminationdue to the use of AI in insurance sector); Karen Hao, Facebook's Ad-serving Algorithm Discriminates by Gender and Race, MIT TECHNOLOGY REVIEW (Apr. 5, 2019), https://www.technologyreview.com/2019/04/05/1175/facebook-algorithm-discriminates-ai-bias/ (arguing how advertising algorithms are inherently discriminatory).

Wired 2019: Supposedly 'Fair' Algorithms Can Perpetuate Discrimination	Fairness; Sikkerhed	"But fairness and accuracy are not necessarily the same thing. For example, when Julia Angwin pointed out in her ProPublica report that risk scores used by the criminal justice system were biased against people of color, the company that sold the algorithmic risk score system argued that its scores were fair because they were accurate. The scores accurately predicted that people of color were more likely to reoffend. This likelihood of reoffense, called the recidivism rate, is the likelihood that someone recommits a crime after being released, and the rate is calculated primarily using arrest data. But this correlation contributes to discrimination, because using arrests as a proxy for recommitting a crime means the algorithm is codifying biases in arrests, such as a police officer bias to arrest more people of color or to patrol more heavily in poor neighborhoods. This risk of recidivism is used to set bail and determine sentencing and parole, and it informs predictive policing systems that direct police to neighborhoods likely to have more crime. There are several obvious problems with this. If you believe the risk scores are accurate in predicting the future outcomes of a certain group of people, then it means it's "fair" that a person is more likely to spend more time in jail simply because they are black. This is actuarially "fair" but clearly not "fair" from a social, moral, or anti-discrimination perspective. The other problem is that there are fewer arrests in rich neighborhoods, not because people there aren't smoking as much pot as in poor neighborhoods but because there is less policing. Obviously, one is more likely to be rearrested if one lives in an overpoliced neighborhood, and that creates a feedback loop—more arrests mean higher recidivism rates. In very much the same way that redlining in minority neighborhoods created a self-fulfilling prophecy of uninsurable communities, overpolicing and predictive policing may be "fair" and "accurate" in the short term, but the long-term effects on communities have been shown to be negative, creating self-fulfilling prophecies of poor, crime-ridden neighborhoods.

		Angwin also showed in a recent ProPublica report that, despite regulations, insurance companies charge minority communities higher premiums than white communities, even when the risks are the same. The Spotlight team at The Boston Globe reported that the household median net worth in the Boston area was $247,500 for whites and $8 for nonimmigrant blacks—the result of redlining and unfair access to housing and financial services. So while redlining for insurance is not legal, when Amazon decides to provide Amazon Prime free same-day shipping to its "best" customers, it's effectively redlining—reinforcing the unfairness of the past in new and increasingly algorithmic ways. Like the insurers, large tech firms and the computer science community also tend to frame "fairness" in a depoliticized, highly technical way involving only mathematics and code, which reinforces a circular logic. AI is trained to use the outcomes of discriminatory practices, like recidivism rates, to justify continuing practices such as incarceration or overpolicing that may contribute to the underlying causes of crime, such as poverty, difficulty getting jobs, or lack of education. We must create a system that requires long-term public accountability and understandability of the effects on society of policies developed using machines. The system should help us understand, rather than obscure, the impact of algorithms on society. We must provide a mechanism for civil society to be informed and engaged in the way in which algorithms are used, optimizations set, and data collected and interpreted. The computer scientists of today are more sophisticated in many ways than the actuaries of yore, and they often sincerely are trying to build algorithms that are fair. The new literature on algorithmic fairness usually doesn't simply equate fairness with accuracy, but instead defines various trade-offs between fairness and accuracy. The problem is that fairness cannot be reduced to a simple self-contained mathematical definition—fairness is dynamic and social and not a statistical issue. It can never be fully achieved and must be constantly audited, adapted, and debated in a democracy. By

		merely relying on historical data and current definitions of fairness, we will lock in the accumulated unfairnesses of the past, and our algorithms and the products they support will always trail the norms, reflecting past norms rather than future ideals and slowing social progress rather than supporting it"
MIT Technology Review 2019: In 2020, let's stop AI ethics-washing and actually do something	Etik	"For all the lip service paid to these issues, many organizations' AI ethics guidelines remain vague and hard to implement. Few companies can show tangible changes to the way AI products and services get evaluated and approved. We're falling into a trap of ethics-washing, where genuine action gets replaced by superficial promises. In the most acute example, Google formed a nominal AI ethics board with no actual veto power over questionable projects, and with a couple of members whose inclusion provoked controversy. A backlash immediately led to its dissolution"
Medium 2018: AI IN 2018: A YEAR IN REVIEW	Sikkerhed; privatliv; fairness	
Leslie/The Alan Turing Institute 2019: Understanding	Lighed	"Because they gain their insights from the existing structures and dynamics of the societies they analyse, datadriven technologies can reproduce, reinforce, and amplify the patterns of marginalisation,

artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector		inequality, and discrimination that exist in these societies. Likewise, because many of the features, metrics, and analytic structures of the models that enable data mining are chosen by their designers, these technologies can potentially replicate their designers' preconceptions and biases. Finally, the data samples used to train and test algorithmic systems can often be insufficiently representative of the populations from which they are drawing inferences. This creates real possibilities of biased and discriminatory outcomes, because the data being fed into the systems is flawed from the start"
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	Ansvarlighed	"When citizens are subject to decisions, predictions, or classifications produced by AI systems, situations may arise where such individuals are unable to hold directly accountable the parties responsible for these outcomes. AI systems automate cognitive functions that were previously attributable exclusively to accountable human agents. This can complicate the designation of responsibility in algorithmically generated outcomes, because the complex and distributed character of the design, production, and implementation processes of AI systems may make it difficult to pinpoint accountable parties. In cases of injury or negative consequence, such an accountability gap may harm the autonomy and violate the rights of the affected individuals"
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	Gennemsigtighed	"Many machine learning models generate their results by operating on high dimensional correlations that are beyond the interpretive capabilities of human scale reasoning. In these cases, the rationale of algorithmically produced outcomes that directly affect decision subjects remains opaque to those subjects. While in some use cases, this lack of explainability may be acceptable, in some applications, where the processed data could harbour traces of discrimination, bias, inequity, or unfairness, the opaqueness of the model may be deeply problematic"

Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	Privatliv	"Threats to privacy are posed by AI systems both as a result of their design and development processes, and as a result of their deployment. As AI projects are anchored in the structuring and processing of data, the development of AI technologies will frequently involve the utilisation of personal data. This data is sometimes captured and extracted without gaining the proper consent of the data subject or is handled in a way that reveals (or places under risk the revelation of) personal information. On the deployment end, AI systems that target, profile, or nudge data subjects without their knowledge or consent could in some circumstances be interpreted as infringing upon their ability to lead a private life in which they are able to intentionally manage the transformative effects of the technologies that influence and shape their development. This sort of privacy invasion can consequently harm a person's more basic right to pursue their goals and life plans free from unchosen influence"
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	Velfærd	"While the capacity of AI systems to curate individual experiences and to personalise digital services holds the promise of vastly improving consumer life and service delivery, this benefit also comes with potential risks. Excessive automation, for example, might reduce the need for human-to-human interaction, while algorithmically enabled hyper-personalisation, by limiting our exposure to worldviews different from ours, might polarise social relationships. Well-ordered and cohesive societies are built on relations of trust empathy, and mutual understanding. As AI technologies become more prevalent, it is important that these relations be preserved"
Leslie/The Alan Turing Institute 2019: Understanding artificial intelligence ethics and safety: A	Sikkerhed	"Irresponsible data management, negligent design and production processes, and questionable deployment practices can, each in their own ways, lead to the implementation and distribution of AI systems that produce unreliable, unsafe, or poor-quality outcomes. These outcomes can do direct damage to the wellbeing of

guide for the responsible design and implementation of AI systems in the public sector		individual persons and the public welfare. They can also undermine public trust in the responsible use of societally beneficial AI technologies, and they can create harmful inefficiencies by virtue of the dedication of limited public resources to inefficient or even detrimental AI technologies"
EDPS Ethics Advisory Group 2018: Towards a digital ethics	Etik	"One of the most encouraging insights recently garnered from industrial practices is that innovation often finds ways to overcome ethical deadlocks and apparently insurmountable value-dilemmas, such as increasing transparency while observing confidentiality, strengthening accountability without breaches of security, or explaining the application of algorithms without reducing the functionality of IT systems"
EDPS Ethics Advisory Group 2018: Towards a digital ethics	Lighed	"Threats to solidarity and empowerment in the digital age are a consequence of the shift to a scored society as outlined in chapter two. They take the form of hyper-individualisation and for a focus on 'real' costs through, for example, behavioural profiling in the context of insurance, the interconnectedness of databases and the use of medical data in the context of employment or in breaches of context-specific rules of confidentiality" (p. 18)
EDPS Ethics Advisory Group 2018: Towards a digital ethics	Lighed	"In the digital age, novel forms of algorithmic discrimination pose a risk to equality of opportunity and to the fundamental right to be protected against digital networks that offer a wealth of often free and accessible information. Unlike traditional economic goods, which obey a law of scarcity, information is multipurpose. The use of digital information for one purpose does not deplete its availability for another. This opens up, on the one hand, a wide array of opportunities for creating and stabilising an economy of sharing based on equality and fairness in a digital society. Yet, on the other hand, the equality of opportunity that is facilitated by the consumption of informational goods by multiple consumers risks creating new inequalities resulting from the fact that some people

		may have the advantage by learning about content before others do and may extract value from it"
EDPS Ethics Advisory Group 2018: Towards a digital ethics	Demokrati; Autonomi	"In the digital age, both the deliberative model of democracy, grounded on citizenship and the notion of the common good are challenged as a basis for the European social contract. Algorithmically processed big data play an increasingly dominant role in informing and guiding individual and social action, in virtually all sectors of business and government. Data-driven governance is often presented as a 'revolutionary' mode of governance emancipated from the yokes of what is assumed to be biased human representation, ambiguous human language, or subjective points-of-view. Personal or anonymous data are the new coordinates of social modelling. Big data rather than institutional or deliberative processes threaten to become the basis on which individuals are classified, evaluated, rewarded or punished. These same categories are used to evaluate the merits and needs of individuals or the opportunities or dangers underlying the lives they lead. In this view of 'data-driven governance', the question arises whether the individual human person as a legal subject has a future and how one can ensure that individuals are not viewed only as temporary data aggregates exploitable on an industrial scale rather than subjects in their own right. Interactions based on algorithmic profiling may exacerbate information imbalances between decision-making governments and companies on the one hand and individuals on the other hand. As a result, 'data-rich' public and private organisations will have greater ethical responsibilities towards citizens and customers. Digital ethics must identify new perspectives, potential and boundaries for dealing with data ethically, by formulating the terms of a proactive approach to ethics, beyond mere legal avoidance measures. As such it will set out the terms of a social innovation that parallels the rapid technological innovation we are experiencing on a daily basis"

EDPS Ethics Advisory Group 2018: Towards a digital ethics	Retfærdighed; Lighed; Fairness	"Criminal investigations are linked to the processing of forensic data and questions of appropriateness and admissibility of data. In criminal justice systems, data-driven algorithmic solutions play a privileged role in the tendency towards performance-oriented management of justice systems. This tendency toward technical management of judicial systems impacts the ecosystem of justice in terms of the presumption of innocence, rules of evidence, processes of justification and the ability to contest judicial decisions, non-discrimination, and equal access to justice. The new horizon of predictive litigation may render law firms more selective in the cases and the individuals they are willing to represent, encouraging advocates to assess the value of sources of evidence by algorithm instead of by human judgment."
EDPS Ethics Advisory Group 2018: Towards a digital ethics	Gennemsigtighed; Ansvarlighed; Sikkerhed	"Data protection faces three interrelated crises of trust: i) individual trust: trust in people, institutions and organisations that deal with personal data is low; ii) institutional trust: transparency and accountability as a condition for keeping track of the reputations of individuals and organisations and trust-building in a society requires access to personal data; and iii) social trust: trust in other members of social groups used to be anchored in personal proximity and physical interaction, which are being increasingly replaced by digital connections. A range of technological fixes to this triple-crisis have appeared on the horizon, though the outcome of their implementation seems unclear: distributed ledger technologies (e.g. blockchain) and peer-to-peer technologies and possibly quantum cryptography could help to solve some of the problems with eroding trust in digital societies. However, blockchains and their functional equivalents give rise to a number of other problems that need to be identified and addressed in due course. In ethical terms, this costly crisis of trust can be addressed by revisiting the terms and qualities of digital communities"

THE CENTRE FOR HUMANITARIAN DATA 2020: GUIDANCE NOTE SERIES DATA RESPONSIBILITY IN HUMANITARIAN ACTION : NOTE #4: HUMANITARIAN DATA ETHICS	Fairness	Needs assessment is required to inform response activities. While the data collected during the assessment enables humanitarian organisations to more effectively target assistance based on need, it also reveals the communities in which households have expressed unfavorable views about local authorities. Local authorities obtain this data and systematically exclude those households from future rounds of assistance, causing direct harm. This compromises the protection of life and health of certain members of the population and undermines the independence of the response as a whole. Although aid organisations are not directly involved in this exclusion, the use of data collected by humanitarians for maleficent purposes may compromise their neutrality
THE CENTRE FOR HUMANITARIAN DATA 2020: GUIDANCE NOTE SERIES DATA RESPONSIBILITY IN HUMANITARIAN ACTION : NOTE #4: HUMANITARIAN DATA ETHICS	Etik	A humanitarian organisation publicly announces a partnership with a private sector technology firm that has also provided technology services to intelligence agencies. Local political groups opposing the presence of the humanitarian organisation use the association with the firm to accuse the humanitarian organisation of spying. Engaging in this political controversy endangers the humanitarian organisation's neutrality.
THE CENTRE FOR HUMANITARIAN DATA 2020: GUIDANCE NOTE SERIES DATA RESPONSIBILITY IN HUMANITARIAN ACTION : NOTE #4: HUMANITARIAN DATA ETHICS	Etik	As a requirement for letting humanitarians into the country, local authorities demand that a 'scientific approach' be taken. They demand that a model be developed to predict humanitarian need and to be involved in the development of that model. The authorities limit the dataset used to train the model in a way that leads to outcomes that overlook critical needs for an oppressed minority. This undermines the protection of life for the minority in question and compromises the neutrality and independence of humanitarian organisations.

THE CENTRE FOR HUMANITARIAN DATA 2020: GUIDANCE NOTE SERIES DATA RESPONSIBILITY IN HUMANITARIAN ACTION : NOTE #4: HUMANITARIAN DATA ETHICS	Etik	An organisation needs to conduct a rapid assessment in order to respond to a displacement in a conflict zone. The data is expected to elicit a negative reaction from the authorities given the nature of the conflict and sensitivity of the humanitarian response in the region in question. While the head of office understands that there is a risk of reprisal toward the staff member if they proceed with the assessment, they also know that they cannot deliver assistance without new data. This situation creates a tension between the need to deliver assistance and the obligation to ensure staff safety.
THE CENTRE FOR HUMANITARIAN DATA 2020: GUIDANCE NOTE SERIES DATA RESPONSIBILITY IN HUMANITARIAN ACTION : NOTE #4: HUMANITARIAN DATA ETHICS	Etik	A model is developed to predict evolving humanitarian need. The model is highly effective in predicting need for most of the population, resulting in a major efficiency gain. However, communities that are less visible in the data are now structurally disadvantaged, creating an algorithmically defined and reinforced population that remains underserved. By obfuscating the needs of a particular group, this undermines the impartiality of the response, calls into question the independence of response organisations, and endangers the life and health of the group(s) underrepresented or missing from the data.
Montreal Declaration Responsible AI u.å.: THE DECLARATION	Etik	The first danger of artificial intelligence development consists in giving the illusion that we can master the future through calculations. Reducing society to a series of numbers and ruling it through algorithmic procedures is an old pipe dream that still drives human ambitions. But when it comes to human affairs, tomorrow rarely resembles today, and numbers cannot determine what has moral value, nor what is socially desirable.

Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission Singapore (PDPC) 2020: Compendium of Use Cases: Practical Illustrations of the Model AI Governance Framework	Ansvarlighed	HUMAN INVOLVEMENT: HOW MUCH IS JUST RIGHT? An online retail store wishes to use AI to fully automate the recommendation of food products to individuals based on their browsing behaviours and purchase history. What is the harm? One possible harm could be recommending products that the customer does not need or want. Is it a serious problem? Wrong product recommendations would not be a serious problem since the customer can still decide whether or not to accept the recommendations. Recommendation: Given the low severity of harm, the human-out-of-the loop approach ["AI makes the final decision without human involvement, e.g. recommendation engines"] could be considered for adoption
Smart Dubai u.å.: ARTIFICIAL INTELLIGENCE PRINCIPLES & ETHICS	Fairness	Following a natural disaster, a government relief agency uses an AI system to detect communities in greatest need by analysing social media data from a range of websites. However those communities where smartphone penetration is lower have less presence on social media, and so are at risk of receiving less attention.
Smart Dubai u.å.: ARTIFICIAL INTELLIGENCE PRINCIPLES & ETHICS	Fairness	An organisation uses an AI tool to automate the pre-screening of candidates for a job opening. It is trained on data from the company's existing employees, the majority of whom are from the same ethnic background. Therefore the system learns to use name and nationality as discriminating factors in filtering job applicants.

Smart Dubai u.å.: ARTIFICIAL INTELLIGENCE PRINCIPLES & ETHICS	Sikkerhed; Fairness	A foreign country has a government service which identifies parents who owe money in child maintenance. The data matching process is often incorrect due to misspelled names or missing data which results in some individuals being incorrectly targeted automatically by the system with the result being a large bill, poor credit ratings and even freezing wages. The recourse for individuals who are incorrectly targeted is time-consuming and not straightforward4. If the potential impact of incorrect decisions had been assessed, mitigation measures such a user-friendly review procedure could have been set up.
Smart Dubai u.å.: ARTIFICIAL INTELLIGENCE PRINCIPLES & ETHICS	Etik	A border camera scanning for predictors of risk may misinterpret a "tic" of an individual with Tourette syndrome as suspicious. These can manifest in a diverse fashion, and should not cause this person to undergo secondary inspection every time they pass through the border5. If the data is updated after the first case is encountered then it would avoid causing inconvenience on subsequent visits.
Smart Dubai u.å.: ARTIFICIAL INTELLIGENCE PRINCIPLES & ETHICS	Sikkerhed; Ansvarlighed	An app that uses AI to assess medical symptoms and has a large user base had to face regulatory scrutiny because of number of complaints from doctors. They warned that the application can miss signs of serious illness.
Smart Dubai u.å.: ARTIFICIAL INTELLIGENCE PRINCIPLES & ETHICS	Gennemsigtighed	A person turned down for a credit card might be told that the algorithm took their credit history, age, and postcode into account, but not learn why their application was rejected1

UNESCO u.å.: Artificial Intelligence: examples of ethical dilemmas	Retfærdighed; godgørenhed; gennemsigtighed; fairness	The use of AI in judicial systems around the world is increasing, creating more ethical questions to explore. AI could presumably evaluate cases and apply justice in a better, faster, and more efficient way than a judge. AI methods can potentially have a huge impact in a wide range of areas, from the legal professions and the judiciary to aiding the decision-making of legislative and administrative public bodies. For example, they can increase the efficiency and accuracy of lawyers in both counselling and litigation, with benefits to lawyers, their clients and society as a whole. Existing software systems for judges can be complemented and enhanced through AI tools to support them in drafting new decisions. This trend towards the ever-increasing use of autonomous systems has been described as the automatization of justice. Some argue that AI could help create a fairer criminal judicial system, in which machines could evaluate and weigh relevant factors better than human, taking advantage of its speed and large data ingestion. AI would therefore make decisions based on informed decisions devoid of any bias and subjectivity. But there are many ethical challenges: - Lack of transparency of AI tools: AI decisions are not always intelligible to humans. - AI is not neutral: AI-based decisions are susceptible to inaccuracies, discriminatory outcomes, embedded or inserted bias. - Surveillance practices for data gathering and privacy of court users. - New concerns for fairness and risk for Human Rights and other fundamental values. So, would you want to be judged by a robot in a court of law? Would you, even if we are not sure how it reaches its conclusions?

UNESCO u.å.: Artificial Intelligence: examples of ethical dilemmas	Lighed; Fairness; Retfærdighed	"Type “greatest leaders of all time” in your favourite search engine and you will probably see a list of the world’s prominent male personalities. How many women do you count? An image search for “school girl” will most probably reveal a page filled with women and girls in all sorts of sexualised costumes. Surprisingly, if you type “school boy”, results will mostly show ordinary young school boys. No men in sexualised costumes or very few. These are examples of gender bias in artificial intelligence, originating from stereotypical representations deeply rooted in our societies. AI-systems deliver biased results. Search-engine technology is not neutral as it processes big data and prioritises results with the most clicks relying both on user preferences and location. Thus, a search engine can become an echo chamber that upholds biases of the real world and further entrenches these prejudices and stereotypes online. How can we ensure more equalised and accurate results? Can we report biased search results? What would or should be the accurate representation of women in search results? Gender bias should be avoided or at the least minimized in the development of algorithms, in the large data sets used for their learning, and in AI use for decision-making. To not replicate stereotypical representations of women in the digital realm, UNESCO wants to address gender bias in AI"
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Lighed; Fairness; Retfærdighed	"AI is likely to have substantial implications for culture and artistic expression. Although still in its infancy, we are beginning to see the first instances of artistic collaboration between intelligent algorithms and human creativity, which might eventually bring important challenges for the rights of artists, the Cultural and Creative Industries (CCI), and the future of heritage" (p. 14) "Creativity, understood as the capacity to produce new and original contentthrough imagination or invention, plays a central role in open, inclusive and pluralistic societies. For this reason, the impact

		of AI on human creativity deserves careful attention. While AI is a powerful tool for creation, it raises important questions about the future of art, the rights and remuneration of artists and the integrity of the creative value chain.56.The case of the 'Next Rembrandt' – in which a brand-new Rembrandt painting was produced using AI and a 3D printer – is a good illustration (Microsoft Europe, 2016). Works of art like this require a new definition of what it means to be an 'author' , in order to do justice to the creative work of both the 'original' author and of the algorithms and technologies that produced the work of art itself. This raises another question: What happens when AI has the capacity to create works of art itself? If a human author is replaced by machines and algorithms, to what extent copyrights can be attributed at all? Can and should an algorithm be recognized as an author, and enjoy the same rights as an artist?57.Although AI is clearly capable of producing 'original' creative works, people are always involved in the development of AI technologies and algorithms, and often in the creation of artworks that serve as the inspiration for AI-generated art. From this perspective, AI can be seen as a new artistic technique, resulting in a new type of art. If we want to preserve the idea of authorship in AI creations, an analysis of the various authors 'behind' each work of art, and their relationships with each other, need to be made. Accordingly, we need to develop new frameworks to differentiate piracy and plagiarism from originality and creativity, and to recognize the value of human creative work in our interactions with AI. These frameworks are needed to avoid the deliberate exploitation of the work and creativity of human beings, and to ensure adequate remuneration and recognition for artists, the integrity of the cultural value chain, and the cultural sector's ability to provide decent jobs" (p. 15)

UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Lighed; Fairness; Retfærdighed	"AI also has a close relation to cultural diversity. While it has the potential to positively impact the cultural and creative industries, not all artists and entrepreneurs have the skills and resources to use AI-based technologies in the creation and distribution of their work. The commercial logic of large platforms may lead to an increased concentration of cultural supply, data and income in the hands of only a few actors, with potential negative implications for the diversity of cultural expressions more generally, including the risk of creating a new creative divide, and an increasing marginalization of developing countries" (p. 15) "Moreover, the algorithms used by media streaming companies such as Spotify and Netflix have a major influence on the selection of music and movies that people enjoy. Because these platforms not only make works of art available, but also suggest works of art for their users to enjoy, it is important that their algorithms are designed in such a way that they do not privilege specific works of art over others by limiting their suggestions to the most dominant works of a particular genre, or to the most popular choices of users and their peers. Other institutions have expressed similar concerns(ARCEP, 2018). Transparency and accountability of these algorithms are essential for ensuring access to diverse cultural expressions and active participation in cultural life
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Demokrati	"The existence of different, sometimes polarized opinions is a regular feature of any open and democratic society that offers a free and open public space. Social media algorithms, however, may exacerbate the polarization of opinions by intensifying and amplifying emotional content via 'likes', - 18 -'shares', 'retweets', auto-completion in search queries, and other forms of online recommendations and engagement, resulting in so-called 'filter bubbles' and 'echo chambers' instead of providing an infrastructure for discussion and debate. Persons sharing the same 'bubble' may be exposed to filtered content of information and in return, the open public space can become characterized with more and more

		homogenized opinion groups which are at the same time more and more polarized to each other" (p. 18)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Etik; ansvarlighed	"Sometimes, the moderation of content can be justified precisely as a means to avoid spreading disinformation and content that incites violence, hatred and discrimination, as well as a means to prevent aggressive personal communication. The filtering may be done by humans, but is often assisted or even automated via AI algorithms. The particular challenge in this case is not just to identify the offending content, but also to avoid the filter being too inclusive and consequently incurring accusations of automated censorship and restriction on legitimate speech. Response to disinformation and 'hate speech' should be based on international freedom of expression standards and in line with UN conventions and declarations on the issue (Article 19, 2018a)" (p. 18)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Etik	"AI can strengthen the free flow of information and journalistic activity, but it can also be used to spread disinformation, which is sometimes referred to using the contested term 'fake news'. Recent examples, such as the Cambridge Analytica affair, have shown that algorithms that were designed to avoid human political bias in deciding which content will appear prominently on social media can be taken advantage of for deliberately promoting the spreading of fabricated, manipulative and divisive content to specific target groups. In some cases, this content may include information fraudulently formatted as news, and may also include content that serves as emotive propaganda. 71.This can have negative effects on norms of civil and informed discussion, on social trust and public debate or even on democratic processes" (p. 17)

UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Ansvarlighed; gennemsigtighed	"Media content production and dissemination increasingly delegate analytical and decision-making authority attributed to sophisticated algorithms. Media organizations increasingly rely on algorithms that analyse user preferences and media consumption patterns (personalization). Applied to journalism, algorithms are then called to analyse specific geographic communities for demographic, social, and political variables in order to produce the most relevant information for these communities, including weather forecasts and sports reports. This practice has the potential to sustain local journalism and newspapers. In this way, AI can help strengthen business models for journalism.77.At the same time, AI-based journalism raises issues of liability, transparency and copyright. Liability can be an issue when it is complicated to determine the fault in algorithm-based reporting, for instance in cases of defamation. Transparency and credibility are issues when consumers do not or cannot realize when content is machine-generated, from which sources it comes and how verified or even false the information is – with current discussions about 'deep fakes' as extreme cases. Copyright is an upcoming issue, since AI-generated content depends ever less on human input, which is reason for some to argue that some form of copyright liability should be attributed to the algorithms themselves" (p. 18)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Etik	"Open educational resources (OER) have been an important addition to the learning landscape with the free availability of high quality lectures and other teaching resources through the internet. The potential of OERs to impact the education of people from across the world is unparalleled, but has yet to be fully realised as the limited completion rates for massive open online courses (MOOCs) demonstrates. The wide variety and depth of resources available has given rise to two problems. Firstly, the problem of finding the right resource for either an individual learner or a teacher wishing to reuse a resource in their own teaching materials. This has led to the second problem of reducing diversity

		through some resources becoming very popular at the expense of other potentially more relevant but less accessible content" (p. 9)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Godgørenhed	"AI has the potential to be beneficial to environmental science through a number of different applications. It can be used to process and interpret data within ecology, systems biology, bioinformatics, space and climate research, thus enhancing scientific understanding of processes and mechanisms. Improved recycling, environmental monitoring and remediation, and more efficient energy consumption can have direct environmental benefits. AI in agriculture and farming can lead to improved crop production (e.g., automated fertilization and irrigation) and animal welfare, and reduced risks from disease, pests, or weather threats. On the other hand, AI could lead to changes in human perceptions of nature, either positively by enhancing human awareness of beauty or independency, or negatively through increased 'instrumentalization' of nature or separation between humans and animals or the environment. 44.For all applications, the potential benefits need to be balanced against the environmental impact of the entire AI and IT production cycle. This includes mining for rare-earth elements and other raw materials, the energy needed to produce and power the machines, and the waste generated during production and at the end of life cycles. Increased AI is likely to add to the growing concerns about the increasing volumes of e-waste and the pressure on rare-earth elements generated by the computing industry. In addition to the environmental and health impacts, e-waste has important socio-political implications, especially related to the export to developing countries and vulnerable populations (Heacock et al., 2015)" (p. 12)

UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Sikkerhed; Gennemsigtighed	"[...] AI can reliably produce impressively accurate predictions based on data sets without giving us any causal or unifying explanation of its predictions. Its algorithms do not work with the same semantic concepts that humans employ to achieve scientific understanding of a phenomenon. This gap between successful predictions on the one hand and satisfactory scientific understanding on the other is likely to play a key role in scientific practice, as well as in decision-making based on AI. 37. This might have implications for trust in science, which is typically based on the scientific method that explains different phenomena in a systematic and transparent way, making its predictions rational and evidence-based. The apparent success of machine learning algorithms to deliver comparable results without such a scientifically justified model could have implications for the public perception and evaluation of science and scientific research. 38. Moreover, research shows that the quality of machine learning depends heavily on the available data used to train the algorithms. But since most AI applications are developed by private companies, there is not always enough transparency about these data, in contrast to the traditional scientific method that warrants the validity of results by requiring replicability, i.e. the possibility to reproduce them by repeating the same experiments" (p. 11)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Etik	"Broadly speaking, social science research aims at finding out the causal structure of personal and social interactions. As most social phenomena are multiply influenced by a number of causal factors, social scientists typically rely on statistical analysis of the relevant empirical data to determine prominent causal factors and the strength of their effects. While doing so, it is crucial to distinguish mere statistical correlations from genuine causal connections. Certainly AI has clear potential to help social scientists navigate huge data sets to come up with plausible causal mechanisms as well as verify the validity of the proposed ones. On the other hand, AI can 'overfit' the data, and put forward 'pseudo' causal relations when there is none. This possibility could lead to social

		controversies especially when the proposed causal relations are ethically sensitive such as suggestions of racial differences of intelligence. Here again we should not accept AI's 'conclusions' automatically without human evaluation"
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Autonomi	"Within the life sciences and medicine in particular, the development of AI technologies has significantly transformed the health care and bioethics landscape over the years. They can bring positive effects, like more precision in robotic surgery, and better care for autistic children, but at the same time, they raise ethical concerns, such as the cost they bring within the context of scarcity of resources in the health care system and the transparency they should bring in order to respect the autonomy of patients [...] From an individual perspective, AI is bringing a new way of dealing with health and medical issues for the lay public. The use of internet sites and the multiplication of mobile phone software applications for self-diagnosis have given people the opportunity to generate health diagnoses without the participation of a health professional. This might have implications for medical authority and for the acceptance of self-medication, including the dangers it entails. It also changes the doctor-patient relationship, and calls for some kind of regulation without hindering innovation and autonomy" (p. 12)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Ansvarlighed	"The possibility exists of the AI-assisted decision-making machine implementing its own attack and kill decisions without human intervention – for example, a fully autonomous weapon. The idea of such a non-human entity having specific agency could radically change our understanding of politics at the widest levels. Moreover, the closeness of potential military uses of AI to its civilian development ('ease of weaponisation') means it is not a discretely bounded category, a characteristic which complicates both the ethics and the regulation of its development and application" (p. 20)

UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Etik	"The speed with which such planning tools could operate would increase the ability to act under rapidly changing situations. One can envisage, for example, the development of algorithmic response to coordinated attack by e.g. drone swarms and other uninhabited assets such as incoming missiles. The speed of AI-enabled response can be seen as an incentive to use it and hence be potentially destabilising. Or indeed disastrous, as past examples of machine warnings being thankfully not being acted on by an intervening human commander have demonstrated. Nevertheless, a State that does not go down this AI response route would be at a major disadvantage, thus encouraging proliferation of the capability" (p. 20)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Etik; Fairness; Retfærdighed; Ansvarlighed	"AI methods can potentially have a huge impact in a wide range of areas, from the legal professions and the judiciary to aiding the decision-making of legislative and administrative public bodies. For example, they can increase the efficiency and accuracy of lawyers in both counselling and litigation, with benefits to lawyers, their clients and society as a whole. Existing software systems for judges can be complemented and enhanced through AI tools to support them in drafting new decisions (CEPEJ, 2018). 49.A key issue in such uses is the nature and interpretation of the results of algorithms, which are not always intelligible to humans1. This issue can be expanded to the wider field of data-driven decision-making. Being able to analyse, process and categorize very large amounts of potentially rapidly-evolving data of very different natures, an AI engine is seen to be capable of proposing – and if allowed, making – decisions in complex situations. Examples of such uses discussed in this report include environmental monitoring, disaster prediction and response, anticipation of social unrest and military battlefield planning.50.The validity of an AI-driven decision however should be treated with caution. Such a decision is not necessarily fair, just, accurate or appropriate. It is susceptible to inaccuracies, discriminatory outcomes, embedded or inserted bias and limitations of the learning process. Not only does a human have a

		much larger 'world view', but he or she also has a tacit knowledge that will outperform AI in critical and complex situations, such as battlefield decisions. Ideally, a decision would be the one a human would make if he or she had been able to process the mountain of data in a reasonable time. However,humans have different capabilities and make decisions based on fundamentally different decision-making architectures, including sensitivity to potential bias [...] Even having a human 'in the loop' to moderate a machine decision may not be sufficient to produce a 'good' decision: as cognitive AI does not make decisions in the same way as humans would, the human would not be equipped with the knowledge and information she or he would need in order to decide if the data-driven action fulfils the human's intentions [...] In some contexts, employing AI as a (either human-assisted or fully autonomous) decision maker might even be seen as a pact with the devil: in order to take advantage of the speed and large data ingestion and categorization capabilities of an AI engine, we will have to give up the ability to influence that decision. Moreover, the effects of such decisions can be profound, especially in conflict situations" (p. 14)
UNESCO 2019: PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE	Lighed; Fairness; Retfærdighed	"A cautionary tale that illustrates some of the problems of using AI to assist decision-making in social contexts is the Allegheny Family Screening Tool (AFST), a predictive model used to forecast child neglect and abuse in Allegheny, Pennsylvania (see https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/07/AFST-Frequently-Asked-Questions.pdf). The tool was put in place with the belief that data-driven decisions would provide the promise of objective, unbiased decisions that would solve the problems of public administration with scarce resources. The Authority that implemented this tool may have been well intentioned. However, recent research has argued that the AFST tool has harmful implications for the population it hoped to serve (Eubanks, 2018b, p.190; Eubanks, 2018a). It oversamples the poor and uses proxies to understand and predict child abuse in a way that inherently disadvantages poor working families. It

		thusexacerbates existing structural discrimination against the poor and has a disproportionately adverse impact on vulnerable communities" (p. 14)
ACCA u.å.: Ethical dilemmas	Privatliv; Gennemsigtighed	https://www.youtube.com/watch?v=SFksSN1inR4&feature=emb_title
ACCA u.å.: Ethical dilemmas	Etik	https://www.youtube.com/watch?v=wGZiI-NAaac&feature=emb_title
ProPublica 2017: What does Facebook consider hate speech?	Etik	"Our analysis of how Facebook implements its hate-speech rules shows that its content reviewers often make different calls on whether to allow or delete items with similar content. To highlight this inconsistency, 3 pairs of posts on the same themes are shown below, along with Facebook's decisions in each case"
ProPublica 2019: Facebook Ads Can Still Discriminate Against Women and Older Workers, Despite a Civil Rights Settlement	Fairness	"Dozens of Companies Are Using Facebook to Exclude Older Workers From Job Ads. Among the companies we found doing it: Amazon, Verizon, UPS and Facebook itself. "It's blatantly unlawful," said one employment law expert" "Facebook Ads Can Still Discriminate Against Women and Older Workers, Despite a Civil Rights Settlement. New research and Facebook's own ad archive show that the company's new system to ensure diverse audiences for housing and employment ads has many of the same problems as its predecessor"

EESC 2016: The ethics of Big Data: Balancing economic benefits and ethical questions of Big Data in the EU policy context		"The creation of digital identities has the obvious advantage of generating the possibility of accessing online contents and all related services through them. The widespread use of digital identities has created fertile ground for the practice of retrieving publicly available information on a person (for example following a job application) from the web, in order to generate insights before actually meeting them. While this process, within boundaries, is accepted as legal, it has the potential to generate discrimination based on the representation of a person as portrayed by their data, as opposed to their real self, in a process known as dictatorship of data where "we are no longer judged on the basis of our actions, but on the basis of what all the data about us indicates our probable actions may be"104 , and personal interaction is placed in a later step after analysis of digital identities"
EESC 2016: The ethics of Big Data: Balancing economic benefits and ethical questions of Big Data in the EU policy context	Etik	"Data processing and analysis are known to be used in providing, during subsequent accesses, personalised results in terms of order of the result pages shown by search engines, marketing offers that are received in our e-mail boxes, advertisements that appear on social networks and other services pages, thus generating a narrower and more personalized version of a user's online experience (the so-called "filter bubble"105). One of the advantages of this extreme personalization is that a user will most likely find what they need in a matter of a few clicks. Nevertheless, this lack of exposure to different items, perspectives and, ultimately, ideas in the long run could represent a strong hindrance factor for creativity and the development of a tolerant attitude by fracturing the reference points necessary for a shared political and social life"

Ingeniøren 2017: Etisk dilemma: Vil du vide alt om dine gendefekter? Sundhedspolitisktidskrift.dk 2019: **Et etisk dilemma: Vævsprøver fra kræftknuder kan afsløre patientens risiko for at udvikle andre sygdomme** Sjællandske Nyheder 2019: Etisk Råd: Deling af sundhedsdata kan give folk uønsket viden Politiken 2016: **Kronik: Store etiske dilemmaer bliver snart rutine**	Autonomi; Sikkerhed	"Etisk dilemma: Vil du vide alt om dine gendefekter? Viden om vores gener får stigende betydning i behandling og forebyggelse af sygdomme, og gensekventering har været anvendt i Danmark i flere år, men netop nu sker der afgørende nyt. Fra nytår vil Rigshospitalet som det første hospital tilbyde kortlægning af hele den genetiske arvemasse som standardbehandling til børn og voksne, der lider af sjældne genetiske sygdomme. Og inden for kort tid forventes sundhedsministeren at fremlægge et lovforslag om etablering af et Nationalt Genom Center, så man centralt opbevarer genetiske data fra patienter. De to ting hænger tæt sammen, for gensekventering er forudsætningen for Danmarks strategi for ' personlig medicin', hvor forebyggelse og behandling i højere grad kan målrettes den enkelte patient [...] sikkerheden omkring data har fået hård kritik. Den går bl. a. på, at staten overtager vores genom, at muligheden for registersamkøring kan blive misbrugt, så data kan bruges til andet end decideret sygdomsbehandling, og at forskere kan bruge data, som de vil forskningsmæssigt, uden at anonymiteten er sikret [...] Ifølge ministeriet skal patienterne stadig give samtykke til genomanalyser, og de kan frabede sig, at deres genom bliver brugt til andet end deres egen behandling. Desuden bliver data pseudonymiseret, så man ikke direkte kan koble data til en bestemt patient. Men hvis en forsker ønsker at gøre brug af dna, kan man finde frem til patienterne og bede om deres samtykke. Sådan er det allerede i dag [...] Lige der har Anne-Marie Gerdes og Det Etiske Råd faktisk en bekymring, for lovforslaget tager ikke højde for, hvordan man sikrer styring af graden af tilbagemeldingen til patienter, hvis forskningen viser øget risiko for en sygdom. »Er det kun dem, vi kan behandle, eller alle, der skal have et svar? Og hvad hvis man ikke ønsker en tilbagemelding? Det er lidt svævende, men bør kunne løses,« siger hun [...] »Det etiske dilemma i forhold til gensekventering handler mest om tilfældighedsfund. Det vil sige, hvis man undersøger en patient for én type sygdom, men samtidig finder ud af, at patienten også er arveligt disponeret for at udvikle andre sygdomme. Men vi har minimeret risikoen for

		tilfældighedsfund, fordi vi lægger filtre ind,« forklarer Anne-Marie Gerdes." (link 1) "Forestil dig, at du er kræftpatient og skal får foretaget en vævsprøve, fordi lægerne vil målrette behandlingen på baggrund af dine genetiske data. De data viser måske, at du har risiko for at udvikle en helt anden, arvelig sygdom. Skal lægen sige det? Vil du vide det?" (link 2) "- Der er fordele og potentiale i at bruge de oplysninger, som patienter selv indsamler. Det kan bruges til at fremme sundheden. - Men samtidig står der en række etiske dilemmaer i kø, der skal løses. Det er vigtigt, at der er åbenhed omkring det, så man ved, hvad ens data bliver brugt til, og det skal være muligt at sige nej til at dele sin data, siger Anne-Marie Axø Gerdes. Samtidig påpeger hun, at der er en uløst udfordring med at sikre, at den data, som borgerne indsamler om sig selv, er af god nok kvalitet, og hvem der adgang til den. Anne-Marie Axø Gerdes frygter et scenarie, hvor en sammenkørsel af ens data kan blive brugt til at lave en sundhedsprofil - for eksempel at man har særlig stor risiko for en bestemt sygdom - som man ikke ønsker at kende til. - Det er forskelligt, hvor meget vi ønsker at vide om vores egen fremtidige sundhed. - Nogle vil gerne vide noget, og andre vil hellere leve i uvidenhed. Det skal vi kunne respektere, siger Anne-Marie Axø Gerdes. Hun mener ikke, at man skal være forpligtet til at dele data, selv om Danmark har et offentligt betalt sundhedsvæsen" (link 3) "Jo flere gener man undersøger, jo større er risikoen for at finde uklare varianter selv i ' kendte' gener. Så man kan risikere at udpege den forkerte genvariant som årsag til patientens

		symptomer og dermed overse den sande årsag til sygdommen med deraf følgende risiko for at vælge forkert behandling og uvirksom medicin [...] Men der er også risiko for, at slægtninge testes for denne genvariant og dermed får beregnet en forkert sygdomsrisiko. Vurderingen af genvarianter er ikke sort-hvid. Der er eksempler på, at genvarianter, der oprindelig blev klassificeret som sygdomsdisponerende, senere har vist sig at være ' uskyldige' genvarianter uden betydning for sygdom. Og omvendt" (link 4)
Berlingske 2017: **DILEMMAER: Tre dataetiske dilemmaer, politikerne bør tage stilling til**	Godgørendhed; Privatliv; Gennemsigtighed; Autonomi	"Helene Ratner, lektor og uddannelsesforsker ved DPU ved Aarhus Universitet (Danmarks Institut for Pædagogik og Uddannelse) opridser følgende tre dilemmaer, som politikerne bør tage stilling til, så vi kan håndtere de mange data:. HVILKE oplysninger skal lærere, skoler og forældre have adgang til? Det kan give mening i et læringsperspektiv at have digitale læringsmidler, som f. eks. tænder for kameraet og bruger eyetracking til at skabe data om elevers måde at læse på eller som registrerer tidsforbrug. Men det skaber samtidig data om, hvor og hvor længe eleven kigger på skærmen, og hvor længe eleven bruger på at læse lektier. Her har vi et dilemma mellem retten til privatliv og data, der kan understøtte læring. Hvordan forholder vi os til den type læringsmidler? Til hvilket formål indsamles data? Hvordan forventes en lærer at fortolke den slags data? Hvordan oplever barnet og forældre det, hvis vi begynder at registrere denne type data? Vil de opleve det som en hjælp eller en øget grad af overvågning? HVORDAN får børn og forældre indsigt i de data, der bliver indsamlet, og de analyser,

		der bliver udarbejdet på baggrund af data? Udgangspunktet er, at man altid bør have mulighed for at få indsigt i data og analyser om sig selv. Men hvordan præsenteres de mange data, så de er brugbare og gennemskuelige for børn og forældre? Og er der undtagelser fra princippet om gennemsigtighed? Med store datasæt er det muligt at lave risikoprofiler for børn i forhold til frafald senere i uddannelsessystemet. Vi har også software, der analyserer børns sociale relationer i klassen. Det kan give læreren mulighed for at gribe præventivt ind og skræddersy indsatser. Men hvordan reagerer et barn på at få at vide, at der kun er få i klassen, som har lyst til at arbejde sammen med det? Eller at det med stor sandsynlighed vil droppe ud? Risikerer vi at demotivere barnet? Risikerer vi, at negative profileringer bliver en selvopfyldende profeti frem for en hjælp? HVORDAN sikrer vi, at de professionelle forstår og er i stand til at bruge data på den rigtige måde (data literacy)? At fortolke data rigtigt kræver viden om og forståelse for data og statistiske usikkerheder. Hvordan tager man højde for, at der altid vil være børn, der ikke passer ind i statistiske modeller? Hvordan klæder vi bedst de professionelle på til at fortolke data i samspil med deres egne observationer og børnenes egne oplevelser? Hvordan sikrer vi, at vi ikke kun samler data ud fra en snæver forståelse af læring, men også får blik for vigtige opgaver som kritisk refleksion, demokratisk dannelse og inklusion?"

Kristeligt Dagblad 2018: Kommuner tager fat på etisk debat om overvågning **Mandag Morgen 2018:** Big data kan hjælpe udsatte, men danskerne er skeptiske	Privatliv	"Kommunerne står med store etiske udfordringer i forhold til, hvordan de skal bruge de stadig mere detaljerede data, de får om den enkelte. En borgmester mener ikke, at kommunerne kan håndtere dilemmaet enkeltvis, fordi de økonomiske interesser vejer for tungt [...] Regeringens ghettoplan har igen skabt debat, nu fordi den indeholder en idé om at give samtlige landets børnefamilier point efter, hvor godt det går med at sende børnene til tandlæge, hvordan det står til med familiens tilknytning til arbejdsmarkedet, om mor og far er skilt og meget andet. Men det er blot en lille forsmag på de dilemmaer, fremtiden bringer med hensyn til overvågning, data og samkøring af data [...] Et eksempel: Hvis det kommunale forsyningsselskab kan se, at " Bjarne på 52 år" fra et belastet boligområde med en løs tilknytning til arbejdsmarkedet bruger strøm, natten igennem, skal den oplysning så videregives til kommunen, så den kan kontakte Bjarne og sige, at " vi tror, du får et problem lige om lidt, og det vil vi gerne gøre noget ved?" [...] Netop Gladsaxe spiller en indirekte rolle i den aktuelle debat om regeringens ghettoplan, fordi idéen om at bruge data til at finde familier med problemer, oprindelig kom fra Gladsaxe i en ansøgning til et frikommuneforsøg. Paradoksalt nok fik kommunen afslag på ansøgningen, og borgmesteren mener ikke, at regeringen nu bruger forslaget, som hendes kommune havde tænkt det. Trine Græse fremhæver, at borgerne skal være med på idéen, og lokalt er hun ikke stødt på modstandmod idéen, fordi det handler om at hjælpe familier, der risikerer at komme i krise. "Det er jo ikke alt, der foregår inden for hjemmets fire vægge, som er en privatsag, for eksempel ikke hvis der er nogle børn, som er i mistrivsel. Dem har vi en forpligtelse til at opspore og gøre noget for. Det er en forpligtelse, vi tager på os, og der vil vi gerne bruge data," sagde Trine Græse. Lidt mere skeptisk er den radikale rådmand i Odense, Susanne Crawley Larsen. Der blev præsenteret et andet eksempel med overvågning, nemlig at en kommune havde udstyret gårdvagter med GPS-udstyr, fordi kameraovervågning havde vist, at der var områder af skolegården, som ikke blev tilset tilstrækkeligt. "Jeg har

		faktisk et problem med, at man overvåger normale mennesker med en normal adfærd. Jeg synes, der skal være en virkelig god grund. Man skal have en rigtig god sag, hvis man gør det i forhold til medarbejdere. Jeg kan også være nervøs for, om det ændrer ens adfærd i den periode, man har GPS'en på sig. Og den eneste anden måde, man kan gøre det på, er ved, at de ikke ved, at de har den på. Det synes jeg er et problem," sagde Susanne Crawley Larsen [...] "Man kan ikke forvente, at de enkelte kommuner har den etiske vinkel. Jeg synes selv, jeg står for holdningspolitik, men når du kommer ind og sidde i et byråd som borgmester, så handler det om økonomi. Man går nogle gange efter en effektivisering - bare tag robotstøvsugeren - men hvad er det så, den efterlader af problemer? Ja, det skal vi da lige love for, at vi har fundet ud af," siger Per Bach Laursen" (link 1) "Tænk, hvis man kunne forhindre, at blot ét barn bliver anbragt uden for hjemmet. Bare ét. Lige præcis den tankegang om at styrke sociale indsatser har gjort, at Gladsaxe Kommune tidligere har ansøgt om at blive frikommune i håbet om at få lov til at samkøre data fra forskellige registre for på den måde at skabe individuelle profiler af borgerne med henblik på mere målrettede sociale indsatser" (link 2)
Prosa 2019: Dataetik handler om dilemmaer i enhver hverdag	Etik	"Det vigtigste dilemma, som panelet fandt frem til, handlede om, hvordan man som medarbejder i en virksomhed eller i det offentlige kan sige fra over for misbrug af data uden at risikere repressalier fra ledelsens side. Selvfølgelig har medarbejdere også et ansvar for at sige fra, hvis borgere eller kunders privatliv krænkes i bestræbelsen på at tjene penge eller effektivisere. Men skal man risikere job og fremtid? Er erstatningerne for en fyring i det tilfælde store nok? Her er ingen lette svar. Både Niels Bertelsen og Simon Tøgern fremførte, at man skulle kollektivisere ansvaret ved at bruge tillidsmandssystemet og dermed fjerne truslen fra den

		enkelte medarbejder. Derfor bliver overenskomster i stigende grad vigtige, og arbejdsgiverne må tegne overenskomster, hvis de vil sikre medarbejderne mulighed for at kritisere brud på dataetikken og opretholde høje dataetiske standarder"
Politiken 2019: IBM-chef står med et dilemma: Tør han bruge Dankortet, når han køber ind?	Fairness	"Han tager udgangspunkt i et supermarked, der har et godt tilbud til kunderne: Hvis de tilmelder sig butikkens rabatordning, vil de få tilsendt skræddersyede tilbud, der er billigere, end de andre kunder i butikken skal betale. Det eneste, de skal gøre til gengæld, er at sætte et flueben, der giver butikken adgang til at høste alle de data, den vil. Systemet er kommet op at køre, og nu er en sælger kommet på besøg for at hjælpe købmanden med at få mest muligt ud af sit nye system. Købmanden: Vi har givet ekstra gode tilbud på vin til mændene, fordi det er dem, der køber mest af det, og de traditionelt har været friskere til at prutte om prisen. Den tradition ville vi gerne føre videre i den digitale verden. Men det har givet en farlig ballade med vrede kvinder, der har opdaget, at deres mænd har fået tilbudt den samme vin til en lavere pris"
Morsø Folkeblad 2019 Sundhedsvæsenet vil følge med på telefonen	Sikkerhed; Privatliv	"For at øge sundheden vil Danske Regioner suge data om borgerne ud af deres telefoner. Men det er ikke alle, der ønsker svar om egen sundhed, advarer Etisk Råd. Det er ikke kun store virksomheder, der er interesserede i at bruge de data, som en smartphone indsamler i lommen på sin bruger. De små personlige databanker kan også få stor betydning i hænderne på det danske sundhedsvæsen. Med dét i sigte har Danske Regioner og Dansk Industri (DI) besluttet sig for at undersøge mulighederne for at få adgang til borgernes personlige data. Det vil give et hidtil uset detaljeret billede af borgerne, vurderer Stephanie Lose, formand for Danske Regioner. -Der er rigtig mange borgere, der går rundt med telefoner, ure eller andre former for digitale devices på sig og opsamler data. -Det kan

		være data om, hvor meget vi går, hvad vores puls er, og nogle kan måle på hjerterytme, siger hun. Sådanne brugerdata vil være en stor fordel, når man skal behandle borgerne og lave forebyggende arbejde, mener Lose. Men ifølge Lars Frelle-Petersen, direktør i DI, kan man også forestille sig, at præcise data bruges i stedet for en fysisk konsultation. På Aalborg Universitet er professor Thomas Ploug ikke i tvivl om potentialet for sundhedsvæsenet ved at skaffe adgang til de personlige data. Men i hænderne på de forkerte kan meget personlige oplysninger også gøre stor skade, lyder det fra forskeren. -Vi kan opnå store fremskridt i behandlingen på baggrund af sådan nogle data. Men der er en række bekymringer, som man skal tage hånd om, og dem skal man tage alvorligt, siger han. Oplysningerne er ekstremt følsomme og vil give et meget præcist billede af borgernes hverdag. I Etisk Råd anerkender formand Anne-Marie Axø Gerdes, at det kan være en fordel for behandlingen i sundhedsvæsenet, hvis lægerne kunne bruge den statistik, som mange danskere indsamler på apps på deres telefoner. Hun ser dog også en række etiske dilemmaer og påpeger, at der er en uløst udfordring med at sikre, at de data, som borgerne indsamler om sig selv, er af god nok kvalitet. Samt med at sikre, hvem der har adgang. Anne-Marie Axø Gerdes frygter et scenarie, hvor en sammenkørsel af ens data kan blive brugt til at lave en sundhedsprofil -for eksempel at man har særlig stor risiko for en bestemt sygdom -som man ikke ønsker at kende til. -Det er forskelligt, hvor meget vi ønsker at vide om vores egen fremtidige sundhed. -Nogle vil gerne vide noget, og andre vil hellere leve i uvidenhed. Det skal vi kunne respektere, siger Anne-Marie Axø Gerdes"
Information 2019: **Deepfake stiller**	Etik	"Techgiganterne er uenige om, hvordan nyere teknologi som deepfakes og andre manipulerede videoer bør håndteres på internettet. Det handler om, hvem der skal have retten til at

techgiganterne over for et nyt dilemma		definere sandt og falsk [...] Hvor skal grænsen gå for manipulerede videoer? Facebook befinder sig i det tilbagevendende dilemma mellem hensynet til brugernes ytringsfrihed og hensynet til at fremme et sikkert og autentisk fællesskab"
Xafis et. al. 2019: An Ethics Framework for Big Data in Health and Research	Privatliv	"The issue of anonymisation has become highly technical and care needs to be taken when making claims about the associated risks or lack thereof, especially because of rapid developments in data science and the different thresholds for considering data 'anonymised' given the various techniques available. In many jurisdictions, there remains a bright regulatory line between 'identifiable4 ' and 'anonymised' data, 'de-identified' data or data that has undergone 'pseudonymisation'. 5 However, the dynamic and multifaceted nature of big data, as well as the variety of data available, has increased the likelihood of privacy threats to data sets that are not readily identifiable. There is increased risk of (re)identification of individuals and/ or a weakening of the security that data masking techniques appear to provide. Three kinds of disclosure risks may lead to the re-identification of an individual despite the masking or de-identification of identifiable data: - identity disclosure—when data is successfully associated with person X; - attribute disclosure —one such disclosure is made when person X is identified as belonging to a particular group, e.g. cancer registry, so there is membership disclosure; and - inferential disclosure—when information about person X can be inferred with high confidence with released data (Templ 2017). Disclosure risks can only ever be completely eliminated if data is not shared at all" (p. 232)
Xafis et. al. 2019: An Ethics Framework for	Autonomi	"A traditional and heavily relied on requirement for including individuals in research or other health-related activities has been informed consent. This standard is essentially an individuals'

Big Data in Health and Research		agreement to assume the potential risk(s) involved in participating in the research or health activity. Heavy reliance on consent is becoming increasingly impracticable in the big data context because data might be linked and used within and across ecosystems that are far removed from the original source of information [...] In such circumstances, it is important to explore alternative ethically acceptable approaches and mechanisms which provide appropriate protections for individuals whose data may be used" (p. 233)
Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	Privatliv	"*There are specific dilemmas that the introduction of the child-tailored online privacy protection regime creates – the 'empowerment versus protection' and the 'individualized versus average child' dilemmas. It concludes that by favouring protection over the empowerment of children, the Regulation risks limiting children in their online opportunities, and by relying on the average child criteria, it fails to consider the evolving capacities and best interests of the child* [...] Achieving an appropriate balance between child protection and participation is not straight-forward"
Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	Sikkerhed; Privatliv	"The collection of data on children is further problematized by the on-sale and sharing of data with third parties, primarily for marketing purposes, but also for alternate uses – known and unknown, none of which are necessarily driven by a directive of the best interests of the child, nor are necessarily open to scrutiny in the public domain. Perhaps the best-known example is that of the controversial 'Smart Barbie' doll produced by Mattell, which led privacy campaigners in 2015 to highlight that recordings of children using voice recognition technology were being sent to third-party companies for processing, potentially revealing his or her intimate thoughts and details (Gibbs, 2015b). All of these data collection methods have the potential to limit the control children have over their information and their public identities. Data mining technologies can create detailed demographic and behavioural

		profiles of children online, raising issues of privacy and intellectual ownership"
Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	Lighed; Fairness; Retfærdighed	"The potential for data mining to give rise to discrimination is a further concern. A literature review regarding the potential for discrimination, arising from big data mining by Barocas (2014), identified three means by which discrimination may occur. Firstly, conscious discrimination may occur, which may be difficult to discern by virtue of the use of algorithms that are premised on underlying factors that may define a particularly vulnerable cohort, such as geographical location or health profile. Secondly, discrimination may result from proportional misrepresentation (under or over representation) of marginalized groups within a particular sample, leading to inaccurate conclusions, rankings and skewed decision making. Finally, discrimination may result from over dependence on specific data sources for decision making - to the exclusion of more verifiable, or nuanced approaches, or the utilization of multiple methods, to allow for triangulation of the data. According to Nissenbaum (2009), the significant driver of discrimination occurs when data is moved out of context, and the contextual integrity of the data is compromised. Pasquale (2014) notes the proliferation of poorly regulated data miners, brokers and resellers, who are providing varied categorizations of persons on a breadth of issues ranging from HIV status, to mental health status, to exposure to sexual abuse. He highlights that these categorical lists raise three ethical issues: first they are frequently inaccurate and almost impossible to verify, second, these lists can and are being inappropriately used for decision making. Finally, people are most likely to be unaware of their presence on these lists"

Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	Lighed; Fairness; Retfærdighed; Gennemsigtighed	"Encompassed within this type of discrimination, is the use of big data for predictive analyses ('predictive analytics'), particularly as it pertains to the identification of 'at risk' youth. Techniques such as predictive risk modelling (PRM) use huge volumes of historical data to evaluate the likelihood of negative events in the future. Using PRM, social service agencies are able to crunch through vast amounts of old case data to provide predictions about which children may face the greatest risk of future harm. The approach – already in widespread use in health care and policing – holds tremendous appeal, especially for cash-strapped social service agencies as it can flag the highest-risk cases for intervention by always-too-few case workers. Initial pilots in countries including New Zealand and the US, have, however, raised concerns about this approach. Issues of data privacy, the underlying drivers of abuse and neglect, and systemic biases, have all been raised by concerned groups, including UNICEF New Zealand (Le Goulven, 2017). Pasquale (2015) notes the use by academics of poorly regulated scoring services to identify potential 'problem students' based on calculations the students cannot access and of which they are unaware. "
Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	Etik	"Tested assessment tools are currently being used or explored in the juvenile justice system, to determine the likely recidivism of juvenile offenders (Judicial Council of California, 2011). However, even the most publically available and validated tools, are providing mixed results. As noted by the Judicial Council of California (2011), given the mixed findings from the validation studies on these instruments, and the limited research currently available, the results from these tools should not be used as the sole determinant of a young person's risk of sexual re-offense (p.4). The use of predictive data in the juvenile justice system is a cause for concern. While the Judicial Council of California (2011) highlighted the need for a cautious and qualified use of tested tools, the reliance on big data and algorithms to determine 'at risk

		youth' have very significant implications for the treatment and sentencing of young people"
Berman & Albright / Unicef 2017: Children and the Data Cycle: Rights and Ethics in a Big Data World	Autonomi	"Furthermore, the value of information no longer resides solely in its primary purpose, but also in potential secondary uses or 'interoperability' of data. In a big data age, even if the notion of informed consent is possible, when the data are first collected, their most innovative secondary uses cannot be imagined. How can organizations provide notice for a purpose that does not yet exist? How can individuals give informed consent to an unknown? In the context of big data, the tried and trusted concept of notice and consent is often either too restrictive to unearth the data's latent value, or too empty to protect an individual's privacy (Nissenbaum, 2013, p.154)"
CIO 2019: Confronting AI's Ethical Dilemmas	Sikkerhed	Executives and other business-level decision makers may have their eyes on amazing new revenue streams that facial recognition can generate. However, a lack of accuracy and lack of standard of handling of such data means those responsible for the collection, storage, analysis and destruction of this data are balancing on a slippery slope of ethical decisions. [...] It's important to bear in mind that, when facial recognition data is collected, ethical data storage decisions must be made, such as - Is long-term storage acceptable for images of children without their parent's consent? - If security footage of an event was recorded, how long should it be kept? - If a retailer identifies a customer with video and verifies it as that person via their credit card transaction, and if the video identifies

		logos on their clothing, is it okay to send targeted advertisements from those logos' organizations to them?
Technical Working Group on Data Collection on Violence against Children : Child Protection Monitoring and Evaluation Reference Group 2012: *Ethical Principles, Dilemmas and Risks in Collecting Data on Violence against Children A review of available literature*	Godgørenhed; Sikkerhed	"Impact on children of participation in research on VAC A key concern for all involved with research on VAC, including researchers, parents, gatekeepers and others, is the impact that participation in research will have on the children involved. Mudaly and Goddard (2009) outline some of the key questions: • Does involvement in child abuse research conflict with ethical principles of beneficence and non-maleficence? • Is it justifiable to include children in abuse research classified as non-therapeutic that has limited, indirect or minor benefits for children? • What are the possible long-term consequences? To be ethical, research must be of sufficient importance, and the benefits must outweigh the risks (King & Churchill, 2000). Foremost among the inherent risks is that participation might cause the child participant distress or trauma (Knight et al., 2000). This may be by way of emotional distress from participation in the research, or harm caused to the child by other people as a consequence of their participation in the research"
The Guardian u.å.: **Facebook bans Rohingya group's posts as minority faces 'ethnic cleansing'**	Etik; Værdighed; Autonomi	"As hundreds of thousands flee a brutal campaign by the Myanmar military, the social media company labels an insurgent group a 'dangerous organization'" "Activists documenting the alleged ethnic cleansing of Rohingya Muslims in Burma are reportedly having their Facebook posts removed and their accounts suspended. Rohingya people who use

Independent 2017: Facebook is 'silencing' Rohingya Muslim reports of 'ethnic cleansing'		Facebook to share information about attacks have called on the company to stop silencing them, the Daily Beast reports. A mass exodus of Rohingya Muslims has sparked allegations of ethnic cleansing, with the UN's High Commissioner of Human Rights calling the operations against them a "textbook example of ethnic cleansing." Attacks by Rohingya insurgents sparked a military response which forced more than 410,000 Rohingya into neighbouring Bangladesh as their villages were burned and hundreds were killed"
The New York Times 2018: 'The Business of War': Google Employees Protest Work for the Pentagon	Etik	"Thousands of Google employees, including dozens of senior engineers, have signed a letter protesting the company's involvement in a Pentagon program that uses artificial intelligence to interpret video imagery and could be used to improve the targeting of drone strikes"
The New York Times 2018: Facebook Security Breach Exposes Accounts of 50 Million Users	Privatliv; Sikkerhed	"Facebook, already facing scrutiny over how it handles the private information of its users, said on Friday that an attack on its computer network had exposed the personal information of nearly 50 million users"
CNN 2019: HUD's new lawsuit against Facebook is a dagger at the heart of the consumer internet	Fairness	"Last week, the US Department of Housing and Urban Development took Facebook and the broader internet industry by surprise and storm with a remarkable allegation: that the company has engaged in discriminatory practices that engendered and perpetuated bias against marginalized classes of the American population -- such as non-Christians, immigrants, and minorities -- by displaying housing ads only to selected audience segments in unfair ways"

Becker's Health IT 2018: **IBM's Watson recommended 'unsafe and incorrect' cancer treatments, STAT report finds**	Sikkerhed	"Internal IBM documents show that its Watson supercomputer often spit out erroneous cancer treatment advice and that company medical specialists and customers identified “multiple examples of unsafe and incorrect treatment recommendations” as IBM was promoting the product to hospitals and physicians around the world"
Independent 2019: Government 'deported 7,000 foreign students after falsely accusing them of cheating in English language tests' Quartz 2018: A flawed algorithm led the UK to deport thousands of students	Sikkerhed	"The government may have mistakenly deported more than 7,000 foreign students after falsely accusing them of cheating in English language tests. Most of the students were not allowed to appeal the Home Office decision; nor were theyt able to obtain evidence against them, or given the opportunity to prove the proficiency in English. Some were detained by immigration officials, lost their jobs, and were left homeless as a result, despite being in the UK legally, the Financial Times reported [...] The firm, English Testing Services, identified 33,725 “invalid” tests taken by students it was confident confident had cheated. The students’ visas were revoked and they were told to leave the country" (Independent 2019) "There’s just one problem: The automated system that ETS used to identify fake test results appears to have been flawed, meaning some of those deportations might not have been justified" (Quartz 2018)
BBC News 2020: George Floyd: Amazon bans police use of facial recognition tech	Lighed; Fairness; Retfærdighed	"Technology giant Amazon has banned the police from using its controversial facial recognition software for a year. It comes after civil rights advocates raised concerns about potential racial bias in surveillance technology. This week IBM also said it would stop offering its facial recognition software for "mass surveillance or racial profiling". The decisions follow growing pressure on firms to respond to the death in police custody of George Floyd. Amazon said the suspension of law enforcement use of its Rekognition

		software was to give US lawmakers the opportunity to enact legislation to regulate how the technology is employed."
Vallor 2018: *An Introduction to Data Ethics*	Fairness; Autonomi	"Rosalina, a promising and hard-working law intern with a mountain of student debt and a young child to feed, is denied a promotion at work that would have given her a livable salary and a stable career path, even though her work record made her the objectively best candidate for the promotion [...] Rosalina's deserved promotion might have been denied because her law firm ranks employees using a poorly-designed predictive HR software package trained on data that reflects previous industry hiring and promotion biases against even the best-qualified women and minorities, thus perpetuating the unjust bias. As a result, especially if other employers in her field use similarly trained software, Rosalina might never achieve the economic security she needs to give her child the best chance for a good life, and her employer and its clients lose out on the promise of the company's best intern"
Weekendavisen 2020: Schaake-doktrinen	Demokrati	"De skader, som Schaake taler om, kommer af, at tech-giganter har fået stadig større indflydelse på vores liv - uden at stater eller myndigheder har fået tilsvarende indflydelse over virksomhederne. Eksempelvis styrer en håndfuld virksomheder de miljøer, som store dele af den demokratiske samtale i dag foregår i, og det er op til dem at træffe de mange svære beslutninger, der følger med. Skal Facebook slette konspirationsteorier om COVID-19? Skal Twitter censurere eller advare, når præsident Trump pynter på sandheden i sine tweets? [...] Reklamevirksomhederne, som Schaake kalder den annoncedrevne forretningsmodel hos virksomheder som Google og Facebook, overtager i øjeblikket flere og flere opgaver fra stater og myndigheder. De bygger databaser over socialt belastede borgere, programmer til at spore sygdomme, cybervåben til forsvaret og overvågningssystemer til politiet - og alle vi andre, altså befolkningen og offentligheden, aner ikke, hvordan den nye

		infrastruktur bliver til. »Virksomhederne har ikke nødvendigvis vores bedste interesser for øje - ikke fordi de er onde kapitalister, men fordi de ikke er drevet af de samme hensyn som offentlige myndigheder og derfor af og til simpelthen træffer dårlige beslutninger på vegne af befolkningen,« siger Schaake"
Mittelstadt et. al. 2016: *The ethics of algorithms:* *Mapping the debate*	Sikkerhed	"When algorithms draw conclusions from the data they process using inferential statistics and/or machine learning techniques, they produce probable6 yet inevitably uncertain knowledge. Statistical learning theory (James et al., 2013) and computational learning theory (Valiant, 1984) are both concerned with the characterisation and quantification of this uncertainty. In addition to this, and as often indicated, statistical methods can help identify significant correlations, but these are rarely considered to be sufficient to posit the existence of a causal connection (Illari and Russo, 2014: Chapter 8), and thus may be insufficient to motivate action on the basis of knowledge of such a connection. The term actionable insight we mentioned earlier can be seen as an explicit recognition of these epistemic limitations. Algorithms are typically deployed in contexts where more reliable techniques are either not available or too costly to implement, and are thus rarely meant to be infallible. Recognising this limitation is important, but should be complemented with an assessment of how the risk of being wrong affects one's epistemic responsibilities (Miller and Record, 2013): for instance, by weakening the justification one has for a conclusion beyond what would be deemed acceptable to justify action in the context at hand [...] Much algorithmic decision-making and data mining relies on inductive knowledge and correlations identified within a dataset. Causality is not established prior to acting upon the evidence produced by the algorithm. The search for causal links is difficult, as correlations established in large, proprietary datasets are frequently not reproducible or falsifiable (cf. Ioannidis, 2005; Lazer

		et al., 2014). Despite this, correlations based on a sufficient volume of data are increasingly seen as sufficiently credible to direct action without first establishing causality (Hildebrandt, 2011; Hildebrandt and Koops, 2010; Mayer-Scho¨nberger and Cukier, 2013; Zarsky, 2016). In this sense data mining and profiling algorithms often need only establish a sufficiently reliable evidence base to drive action, referred to here as actionable insights. Acting on correlations can be doubly problematic.10 Spurious correlations may be discovered rather than genuine causal knowledge. In predictive analytics correlations are doubly uncertain (Ananny, 2016). Even if strong correlations or causal knowledge are found, this knowledge may only concern populations while actions are directed towards individuals (Illari and Russo, 2014)"
Mittelstadt et. al. 2016: *The ethics of algorithms:* *Mapping the debate*	Gennemsigtighed	"When data are used as (or processed to produce) evidence for a conclusion, it is reasonable to expect that the connection between the data and the conclusion should be accessible (i.e. intelligible as well as open to scrutiny and perhaps even critique).7 When the connection is not obvious, this expectation can be satisfied by better access as well as by additional explanations. Given how algorithms operate, these requirements are not automatically satisfied. A lack of knowledge regarding the data being used (e.g. relating to their scope, provenance and quality), but more importantly also the inherent difficulty in the interpretation of how each of the many data-points used by a machine-learning algorithm contribute to the conclusion it generates, cause practical as well as principled limitations (Miller and Record, 2013) [...] The scrutability of evidence, evaluated in terms of the transparency or opacity of algorithms, proved a major concern in the reviewed literature. Transparency is generally desired because algorithms that are poorly predictable or explainable are difficult to control, monitor and correct (Tutt, 2016)

		[...] Burrell (2016) and Schermer (2011) argue that the opacity of machine learning algorithms inhibits oversight. Algorithms ‘‘are opaque in the sense that if one is a recipient of the output of the algorithm (the classification decision), rarely does one have any concrete sense of how or why a particular classification has been arrived at from inputs’’ (Burrell, 2016: 1). Both the inputs (data about humans) and outputs (classifications) can be unknown and unknowable. Opacity in machine learning algorithms is a product of the highdimensionality of data, complex code and changeable decision-making logic (Burrell, 2016). Matthias (2004: 179) suggests that machine learning can produce outputs for which ‘‘the human trainer himself is unable to provide an algorithmic representation.’’ Algorithms can only be considered explainable to the degree that a human can articulate the trained model or rationale of a particular decision, for instance by explaining the (quantified) influence of particular inputs or attributes (Datta et al., 2016)"
Mittelstadt et. al. 2016: *The ethics of algorithms:* *Mapping the debate*	Lighed; Fairness; Retfærdighed	"Algorithms process data and are therefore subject to a limitation shared by all types of data-processing, Figure 1. Six types of ethical concerns raised by algorithms. namely that the output can never exceed the input. 4 Big Data & Society While Shannon’s mathematical theory of communication (Shannon and Weaver, 1998), and especially some of his information-inequalities, give a formally precise account of this fact, the informal ‘garbage in, garbage out’ principle clearly illustrates what is at stake here, namely that conclusions can only be as reliable (but also as neutral) as the data they are based on. Evaluations of the neutrality of the process, and by connection whether the evidence produced is misguided, are of course observer-dependent [...] The automation of human decision-making is often justified by an alleged lack of bias in algorithms (Bozdag, 2013; Naik and Bhide,

		2014). This belief is unsustainable, as shown by prior work demonstrating the normativity of information technologies in general and algorithm development in particular14 (e.g. Bozdag, 2013; Friedman and Nissenbaum, 1996; Kraemer et al., 2011; Macnish, 2012; Newell and Marabelli, 2015: 6; Tene and Polonetsky, 2013b). Much of the reviewed literature addresses how bias manifests in algorithms and the evidence they produce [...] Algorithms inevitably make biased decisions. An algorithm's design and functionality reflects the values of its designer and intended uses, if only to the extent that a particular design is preferred as the best or most efficient option. Development is not a neutral, linear path; there is no objectively correct choice at any given stage of development, but many possible choices (Johnson, 2006). As a result, ''the values of the author [of an algorithm], wittingly or not, are frozen into the code, effectively institutionalising those values'' (Macnish, 2012: 158). It is difficult to detect latent bias in algorithms and the models they produce when encountered in isolation of the algorithm's development history (Friedman and Nissenbaum, 1996; Hildebrandt, 2011; Morek, 2006). Friedman and Nissenbaum (1996) argue that bias can arise from (1) pre-existing social values found in the ''social institutions, practices and attitudes'' from which the technology emerges, (2) technical constraints and (3) emergent aspects of a context of use. Social biases can be embedded in system design purposefully by individual designers, seen for instance in manual adjustments to search engine indexes and ranking criteria (Goldman, 2006). Social bias can also be unintentional, a subtle reflection of broader cultural or organisational values. For example, machine learning algorithms trained from human-tagged data inadvertently learn to reflect biases of the taggers (Diakopoulos, 2015)" (p. 7)

Mittelstadt et. al. 2016: *The ethics of algorithms:* *Mapping the debate*	Lighed; Fairness; Retfærdighed	"The three epistemic concerns detailed thus far address the quality of evidence produced by an algorithm that motivates a particular action. However, ethical evaluation of algorithms can also focus solely on the action itself. Actions driven by algorithms can be assessed according to numerous ethical criteria and principles, which we generically refer to here as the observerdependent 'fairness' of the action and its effects. An action can be found discriminatory, for example, solely from its effect on a protected class of people, even if made on the basis of conclusive, scrutable and well-founded evidence" "Much of the reviewed literature also addresses how discrimination results from biased evidence and decisionmaking.15 Profiling by algorithms, broadly defined ''as the construction or inference of patterns by means of data mining and ... the application of the ensuing profiles to people whose data match with them'' (Hildebrandt and Koops, 2010: 431), is frequently cited as a source of discrimination. Profiling algorithms identify correlations and make predictions about behaviour at a group-level, albeit with groups (or profiles) that are constantly changing and re-defined by the algorithm (Zarsky, 2013). Whether dynamic or static, the individual is comprehended based on connections with others identified by the algorithm, rather than actual behaviour (Newell and Marabelli, 2015: 5). Individuals' choices are structured according to information about the group (Danna and Gandy, 2002: 382). Profiling can inadvertently create an evidencebase that leads to discrimination (Vries, 2010)"

Mittelstadt et. al. 2016: *The ethics of algorithms:* *Mapping the debate*	Autonomi	The ethical challenges posed by the spreading use of algorithms cannot always be retraced to clear cases of epistemic or ethical failures, for some of the effects of the reliance on algorithmic data-processing and (semi-) autonomous decision-making can be questionable and yet appear ethically neutral because they do not seem to cause any obvious harm. This is because algorithms can affect how we conceptualise the world, and modify its social and political organisation (cf. Floridi, 2014). Algorithmic activities, like profiling, reontologise the world by understanding and conceptualising it in new, unexpected ways, and triggering and motivating actions based on the insights it generates. [...] Value-laden decisions made by algorithms can also pose a threat to the autonomy of data subjects. The reviewed literature in particular connects personalisation algorithms to these threats. Personalisation can be defined as the construction of choice architectures which are not the same across a sample (Tene and Polonetsky, 2013a). Similar to explicitly persuasive technologies, algorithms can nudge the behaviour of data subjects and human decision-makers by filtering information (Ananny, 2016). Different content, information, prices, etc. are offered to groups or classes of people within a population according to a particular attribute, e.g. the ability to pay. [...] Algorithms are also driving a transformation of notions of privacy. Responses to discrimination, de-individualisation and the threats of opaque decision-making for data subjects' agency often appeal to informational privacy (Schermer, 2011), or the right of data subjects to ''shield personal data from third parties.'' Informational privacy concerns the capacity of an individual to control information about herself (Van Wel Mittelstadt et al. 9 and Royakkers, 2004), and the effort required by third parties to obtain this information.

Mittelstadt et. al. 2016: *The ethics of algorithms:* *Mapping the debate*	Ansvarlighed	Algorithms are software-artefacts used in data-processing, and as such inherit the ethical challenges associated with the design and availability of new technologies and those associated with the manipulation of large volumes of personal and other data. This implies that harm caused by algorithmic activity is hard to debug (i.e. to detect the harm and find its cause), but also that it is rarely straightforward to identify who should be held responsible for the harm caused.8 When a problem is identified addressing any or all of the five preceding kinds, ethical assessment requires both the cause and responsibility for the harm to be traced. Thanks to this map (Figure 1), we are now able to distinguish epistemological, strictly ethical and traceability types in descriptions of ethical problems with algorithms. The map is thus intended as a tool to organise a widely dispersed academic discourse addressing a diversity of technologies united by their reliance on algorithms. To assess the utility of the map, and to observe how each of these kinds of concerns manifests in ethical problems already observed in algorithms, a systematic review of academic literature was carried out.9 The following sections (4 to 10) describe how ethical issues and concepts are treated in the literature explicitly discussing the ethical aspects of algorithms. [...] When a technology fails, blame and sanctions must be apportioned. One or more of the technology's designer (or developer), manufacturer or user are typically held accountable. Designers and users of algorithms are typically blamed when problems arise (Kraemer et al., 2011: 251). Blame can only be justifiably attributed when the actor has some degree of control (Matthias, 2004) and intentionality in carrying out the action. Traditionally, computer programmers have had ''control of the behaviour of the machine in every detail'' insofar as they can explain its design and function to a third party (Matthias, 2004). This traditional conception of responsibility in software design assumes the programmer can reflect on the technology's likely

		effects and potential for malfunctioning (Floridi et al., 2014), and make design choices to choose the most desirable outcomes according to the functional specification (Matthias, 2004). With that said, programmers may only retain control in principle due to the complexity and volume of code (Sandvig et al., 2014), and the use of external libraries often treated by the programmer as 'black boxes' (cf. Note 7)
Markedsføring.dk 2020: Danske brands fanget i Amazon-dilemma	Etik	"Den amerikanske ehandelsgigant Amazon er rykket nærmere på Danmark med lanceringen af et svensk site, og det stiller mange danske brands og netkøbmænd i stort dilemma. – Det består i, om de skal gå efter de eksport-muligheder, kendskab og større volumen, som en tilstedeværelse på Amazon kan give, eller sige nej tak og i stedet satse på egen webbutik og de fordele, det giver i forhold til loyalitetsopbygning og markedsføring, siger Dorte Wimmer, director for Retail Institute Scandinavia. – De danske brands og webshops skal gøre op med sig selv, om de gerne vil have en del af den ehandelskage, som Amazon kan give dem, og holde dette op mod de ting, de må give slip på – ikke mindst på mindst datasiden – hvis de vælger at satse på Amazon-platformen, siger hun"