DATAETIK

Principper og pejlemærker for
virksomheder, myndigheder
& organisationer

den uafhængige tænkehandletank

DATAETIK – Principper og pejlemærker for virksomheder, myndigheder og organisationer
1. udgave 2018

Forfattere: Pernille Tranberg, Gry Hasselbalch, Birgitte Kofod Olsen & Catrine Søndergaard Byrne
Grafisk opsætning: Spintype.com
Omslag: Paws Fabrik
Printed by: AKAPRINT A/S
Publiceret med: Spintype.com

Isbn print: 9788771920444
Isbn pdf: 9788771920451
Isbn epub: 9788771920468

INDHOLDSFORTEGNELSE

KAPITEL 1

INTRODUKTION

Dataetik er teknologiens krumtap. I hvert fald for os, som længe har arbejdet med emnet. I 2015 etablerede vi DataEthics.eu, en europæisk uafhængig tænkehandletank, der arbejder konstruktivt for at fremme dataetisk praksis. I denne håndbog præsenterer vi DataEthics.eu's dataetiske principper med uddybende spørgeramme og FAQ. Den vejleder omkring processen med at integrere dataetik i den daglige behandling af persondata. Principperne kan bruges frit, hvis det tydeligt fremgår, at de er udviklet af DataEthics.eu, og der linkes til dataethics.eu/dataetiske-principper/

Ingen er perfekte. Alt er i beta – også inden for dataetik. Ligesom ingen gør det perfekt, når det gælder fx miljøansvarlighed, er der ingen, der er fuldkomne, når det gælder dataetik. Det handler om at få gang i processen, om at få de dataetiske tanker ind bag nethinden, begynde at praktisere det og blive bedre til det – skridt for skridt.

Pernille, Gry, Birgitte og Catrine, stiftere af DataEthics.eu
September 2018

KAPITEL 2

DEFINITION PÅ DATAETIK

Dataetik er ansvarlig og bæredygtig brug af data. Det er det "rigtige" at gøre i forhold til mennesker og samfund. Dataprocesser bør være designet som bæredygtige løsninger – dvs. først og fremmest til gavn for mennesker.

Dataetik handler om at opfylde de principper og værdier, som menneskerettighederne og persondatalovgivningen bygger på. Det handler om ærlig og ægte gennemsigtighed i datahåndteringen. Om aktivt at udvikle privacy by design og privacy-fremmende produkter og strukturer. Om at behandle andres personlige oplysninger

som man selv ønsker ens egne – eller ens børns – skal behandles.

Dataetik er skridtet videre end efterlevelse af persondatalovgivningen: Alle dataprocesser respekterer derfor som minimum de krav, der er opstillet i EU's Persondataforordning (GDPR), EU's Charter om grundlæggende rettigheder samt den Europæiske Menneskerettighedskonvention.

KAPITEL 3

DATAETISKE PRINCIPPER

MENNESKET I CENTRUM

Menneskets interesser har altid forrang for institutionelle og kommercielle interesser. Mennesket er ikke en dataproces eller et stykke software, men unikt med menneskelig empati, vilje, uforudsigelighed, intuition og kreativitet, og mennesket har en højere status end maskinen. Personen er i centrum og har den primære gavn af databehandlingen.

INDIVIDUEL DATAKONTROL

Mennesket har individuel datakontrol og handlekraft. Personens selvbestemmelse prioriteres i alle dataprocesser og personen tager aktiv part i de data, der registreres om dem. Det er det enkelte menneske, der har den primære kontrol over, hvad deres data bruges til og i hvilke sammenhænge, samt over hvordan deres data aktiveres.

GENNEMSKUELIGHED

Databehandling og automatiserede beslutninger skal give mening for det enkelte menneske. De skal være transparente og skal kunne forklares. Formål med og interesser i databehandlingen skal være gennemskuelige for mennesket i forhold til at forstå risici, samt sociale, etiske og samfundsmæssige konsekvenser.

ANSVARLIGHED

Ansvarlighed er en organisations bevidste, saglige og systematiske brug og beskyttelse af persondata. Ansvar og medansvar er med i alle led i en databehandling, og der gøres en aktiv indsats for at mindske risici for individet samt at inddæmme sociale og etiske konsekvenser. Bæredygtig persondatabehandling er indlejret i hele organisationen og sikrer etisk ansvarlighed på både kort, mellem og lang sigt. En organisations ansvarlighed bør også gælde for underleverandører og samarbejdspartnere.

LIGEVÆRDIGHED

Demokratisk databehandling tager udgangspunkt i, at datasystemer er med til at e bevare, reproducere og skabe magtfordelingen i samfundet. I en databehandling skal der tages særlige hensyn til sårbare mennesker, som eksempelvis

på grund af deres økonomiske, sociale og sundhedsmæssige forhold er særligt udsatte for profilering, der kan have negativ effekt på deres selvbestemmelse og kontrol, eller udsætte dem for diskrimination eller stigmatisering. Hensyn til sårbare mennesker er også at arbejde aktivt med at mindske bias i udviklingen af selvlærende algoritmer.

KAPITEL 4

DATAETISK SPØRGERAMME

Tjek jeres organisations dataetik. Med nedenstående spørgsmål kan I komme rundt i de dataetiske dilemmaer. Brug evt. jeres svar som grundlag for at udarbejde dataetiske retgningslinjer.

MENNESKET I CENTRUM

- *Er jeres databehandling tilrettelagt med udgangspunkt i, at I låner data af brugerne?*
- *Sikrer I, at brugerens rettigheder bliver prioriteret frem for kommercielle eller institutionelle interesser?*

- *Sikrer I, at det først og fremmest er brugerne, der får værdi ud af deres egne data - ikke kun organisationen?*
- *Benytter I privacy by design principper, og kan I beskrive dem klart og gennemskueligt?*

INDIVIDUEL DATAKONTROL

On-device processing

- *Sikrer I, at brugernes data så vidt muligt behandles direkte på brugernes egne devices?*
- *Når der er behov for behandling af data på andet end brugernes egne devices, eksempelvis egen server eller cloud-løsning, indsamles data så de ikke kan knyttes til en identificerbar person?*

Profilering

- *Bruger I profilering? Og giver I brugeren mulighed for at influere og bestemme de værdier, regler og input, der ligger til grund for profileringen?*

Forudsigelser

- *Bruger I data til at forudsige adfærd på individniveau eller kun til mønstre?*

GENNEMSKUELIGHED

Datalagring

- *I hvilket land opbevares data?*
- *Hvor har udbyderen af opbevaringsløsninger hovedsæde?*
- *Går transmissionen af data gennem lande uden for EU?*

Kunstig intelligens

- *Bruger I machine learning/kunstig intelligens? Hvis ja, kan I forklare jeres algoritme - kriterierne og parametrene?*

Adfærdsdesign

- *Bruger I personlig data til at påvirke adfærd?*
- *Sikrer I, at det er gennemskueligt for brugerne, at deres adfærd bliver påvirket?*
- *Sikrer I, at designet ikke skaber afhængighed og dermed fjerner personens selvbestemmelse og handlekraft?*

Open Source

- *Opererer I med åbne økosystemer (open source software), så andre kan bruge det og evt. arbejde videre med det?*

ANSVARLIGHED

Anonymitet

- *Hvornår anonymiserer I persondata?*
- *Anvender I end-to-end-kryptering af data?*
- *Minimerer I brugen af metadata, og forklarer I hvordan?*

Zero-knowledge

- *Anvender I zero-knowledge som design- og behandlingsprincip?*

Salg af data

- *Sælger I data til tredjepart?*
- *Sælger I data som personhenførbare data?*
- *Sælger I data som mønstre i aggregeret form?*
- *Hvis I sælger data, sikrer I så, at data er fuldt anonymiserede og kun beskriver mønstre, ikke individer?*

Datadeling

- *Bruger I tredjepart cookies?*
- *Omfatter disse SoMe (social media)-cookies og SoMe logon?*
- *Benytter I Google Analytics eller ligneden tracking redskaber?*
- *Er jeres brugere fuldt ud bevidste om, at jeres cookie-brug betyder, at I deler data om dem med tredjepart, og indforstået med det?*

Databerigelse

- *Beriger I data med eksterne data, fx fra sociale medier eller web-scraping?*
- *Sker denne berigelse på foranledning af eller i samarbejde med jeres brugere?*

Organisatorisk forankring

- *Har I en person eller enhed, der er ansvarlig for etisk håndtering af data?*

- *Hvordan er arbejdet med dataetik organisatorisk forankret?*
- *Hvordan sikrer I, at jeres dataetiske retningslinjer overholdes?*

Ekstern kontrol

- *Kan databehandlingen blive auditeret af uafhængig tredjepart?*
- *Stiller I krav til og kontrollerer I jeres underleverandører og samarbejdspartneres dataetik?*

LIGEVÆRDIGHED

Offentlige platforme

- *Har I dialog med jeres brugere på en offentlig platform?*
- *Har I retningslinjer for brugen af platformen?*

- *Modererer I platformen med henblik på at fjerne følsomme persondata?*
- *Hvis I udbyder tjenester til børn e, sikrer I så forældresamtykke?*

Genbrug

- *Bruges data til at udvikle eller træne en algoritme?*
- *Sikrer I, at brugen af data ikke fører til diskrimination?*
- *Sikrer I, at brugen af data ikke fører til udstilling af sårbarheder hos individer?*

Kunstig intelligens

- *Sikrer I, at brugen af kunstig intelligens/machine learning er til gavn for individet og ikke medfører fysisk, psykisk, social eller økonomisk skade for individet?*

KAPITEL 5

FAQ PÅ DATAETIK

Nedenfor følger en række ofte stillede spørgsmål, når det gælder DataEthics.eu's dataetiske principper og vores svar i alfabetisk rækkefølge.

Adfærdsdesign

Hvad er dataetisk adfærdsdesign?

Brug af persondata til at påvirke brugeres adfærd kan være manipulerende, såfremt brugeren ikke er i centrum, men designet primært er udviklet for at skabe afhængighed, flere brugere, mere brug og fremme salget. Dataetisk adfærdsdesign er gennemskueligt for brugeren, ikke-diskriminerende og ikke afhængighedsska-

bende. Det enkelte menneske skal kunne bevare sin selvbestemmelse og handlekraft.

Aktiv part

Hvad betyder det at være aktiv part?

Hvis en læge skriver noget om dig i en patientjournal, eller en lærer registrerer noget om dit barn eller dig som far, som du mener er unuanceret, så kan du bidrage med flere oplysninger, som er synlige for alle, der har adgang til dine data. Et dataetisk forsikringsselskab giver dig også adgang til at kommentere på de konklusioner, de har draget baseret på din data.

Aktivering af data

Hvad vil det sige at aktivere sine data?

GDPR giver det enkelte menneske ret til at kontrollere egne data og til "portabilitet", dvs. at få sine data udleveret i et brugbart format, så man fx kan bruge dem igen hos en konkurrent (fx hvis du vil have din historik med over dit elfor-

brug). Individuel datakontrol er dog ikke nok på længere sigt. Individer vil også fremadrettet stille krav om at kunne aktivere egne data og sætte dem i spil med det formål at kunne berige egen økonomi, sundhed og hverdag. Dette vil også være til gavn for den virksomhed eller institution, der stiller nye tjenester til rådighed, hvor individet kan aktivere egne data.

Anonymisering

Hvad er etisk ansvarlig anonymisering?

Pseudonymisering betyder, at man ikke direkte kan se, hvem oplysningerne vedrører, men at man har stadig mulighed for at genskabe identiteten på de personer. Anonymisering er skridtet videre. Ingen bør ved anonymisering kunne genskabe identiteten. Både ved pseudonymisering og anonymisering af data er det vigtigt at kunne dokumentere dette og tillade en tredjepart at se ind i maskinrummet for at verificere, evt. certificere, dette. Der er ikke mange eksterne tredje-

parter, der tilbyder at gøre dette i dag, men det bliver en vigtig indsats fremadrettet, jf. "Ekstern kontrol".

Bias

Hvad er bias i design?

Bias er indbyggede fordomme og negative stereotyper. Bias kan forekomme i træningsdata, dvs. de historiske data, man bruger til at udvikle en selvlærende algoritme. Bias kan også forekomme i designet af en algoritme, som kan kategorisere og stemple mennesker på en måde, der diskriminerer mellem fx befolkningsgrupper. Dette kan mindskes ved manuel sortering og oprydning i data, og ved at sikre, at algoritmen kan forklares og kan efterses af en uafhængig instans. Fx var vinderne i den første skønhedskonkurrence, der blev bedømt af en selvlærende algoritme, stort set alle hvide mennesker, fordi algoritmen var blevet trænet på flest billeder af hvide mennesker. Der manglede således

flere billeder af sorte mennesker i selve træningen af algoritmen, der bedømte indsendte billeder.

Databerigelse

Hvad er web-scraping, og kan det gøres etisk ansvarligt?

Man kan berige sine data med web-scraping fra websites, herunder sociale medier (alt det, der er åbent og offentligt for alle). Men det er kontroversielt, for selv om data er tilgængelige for alle og lovlige at bruge, er det ikke sikkert, at brugerne synes, det er etisk forsvarligt. Derfor skal du sikre dig, at det sker på foranledning af og med brugernes informerede samtykke.

Datadeling

Hvornår er det uetisk at bruge tredjeparts cookies?

Hvis en virksomhed eller organisation har med børn og andre sårbare at gøre, er det ikke dataetisk at tillade tredjeparts cookies på ens hjem-

meside, da det er lig med deling af identificerbare følsomme data med tredjepart. Hvis man har med sundhedsdata eller data omkring politiske tilhørsforhold, seksuel eller religiøs orientering eller andre følsomme data at gøre, er det heller ikke dataetisk at tillade tredjeparts cookies på ens hjemmeside. Offentlige instanser, der deler data om borgernes adfærd via tredjeparts cookies, herunder SoMe-cookies, er heller ikke dataetiske. De færreste forbrugere og borgere forstår datadelingen og opfatter den som skjult, selv om de har sagt ja til den via en pop-up. Det kan dermed være lovligt, men det overtræder flertallets creepines grænse og må derfor anses som uetisk.

Datalagring

Hvornår er datalagring etisk problematisk?

Det kan sagtens være lovligt at lagre data i et land uden for EU. Men ikke alt, der er lovligt, er etisk. Det kan for eksempel diskuteres, om

det overhovedet er etisk forsvarligt at lagre sine data i en virksomhed med hovedsæde i et land, der praktiserer og tillader datadiktatur eller datamonopoler. Det mener vi ikke. Lagring af data i egne servere, hvor man altid har kontrol over data eller i en cloud-provider med hovedsæde i EU/EØS, kan derimod godt betragtes som dataetisk.

Ekstern kontrol

Hvorfor uafhængig auditering?

Det er vigtigt – i hvert fald fremadrettet – at jeres databehandling kan tåle at blive gennemgået og verificeret af en uafhængig ekstern auditør. Ligesom med miljøet, børnearbejde og IT-sikkerhed får flere og flere brugere behov for at vide, at det I siger, også er det, I gør. I dag findes kun få troværdige ordninger, som verificerer eller certificerer, såsom ISO og Europrise, men i kølvandet på GDPR har EU varslet en decideret europæisk privacy-mærknings-

ordning, ligesom der uden tvivl også kommer flere andre typer mærkningsordninger.

End-to-end kryptering

Hvad er kryptering?

En ting er at kryptere trafikken, så ingen kan opsnappe data, der er i transit. En anden ting er såkaldt end-to-end-kryptering, hvor ingen andre end afsender og modtager kan se indholdet – heller ikke den virksomhed, der ejer platformen, hvor kommunikationen finder sted. Især det sidste er blevet et godt dataetisk argument.

Forklarlighed

Hvad er forklarlighed?

Algoritmer skal kunne forklares, så mennesket forstår det. De skal ikke blot give basal information om databehandlingen, men det skal være dokumenteret og forklaret, hvordan en given algoritmisk beslutning er truffet; hvilke krite-

rier og parametre der ligger til grund for fx en kreditvurdering, forsikringspræmie eller tildeling af social ydelse.

Forudsigelser

Hvad er dataetiske forudsigelser?

Det kan være okay at udarbejde individuelle forudsigelser med fx personaliseret medicin og behandling. Det dataetiske spørgsmål er, hvorvidt individet har indsigt og mulighed for indsigelse samt til at vælge til og fra.

Kunstig intelligens (AI)

Hvordan håndterer vi kunstig intelligens bedst?

Det gør vi kun ved at sikre menneskelig kontrol. DTU Compute har formuleret en række meget fine Safe AI-principper:

- **Safe AI er sikker** og har bestået test og verifikation og er robust over for systematiske og velinformerede angreb.

- **Safe AI er selvbevidst** fordi AI forstår sin egen rolle og usikkerhed og kan fx afslå at handle
- **Safe AI kan holde på en hemmelighed** og beskytte privatliv. Privacy by design er indbygget
- **Safe AI har veldefinerede værdier** og er renset for stereotyper og bias samt forstår emotioner
- **Safe AI har sociale kompetencer,** fordi AI forstår sociale relationer og forstår brugerens viden og kompetencer
- **Safe AI forstår magt** dvs forstår data og handlingers kontekst og konsekvenser
- **Safe AI er dokumenteret** transparent og kommunikerende, right to explanation
- **Safe AI er open source:** metoder, koder og testresultater er tilgængelige for alle

Kilde: Professor Lars Kai Hansen, DTU Compute

Metadata

Hvad er metadata?

Metadata er "data om data" eller en slags varedeklaration for data, der kan give information om datasæt og -tjenester. Det kan eksempelvis være data om, hvem der har sendt en email til hvem og hvornår. Metadata fortæller altså intet om, hvad der bliver skrevet i e-mailen. Der har været diskussioner om, hvorvidt metadata er personoplysninger, dvs. om det er muligt at identificere en person på baggrund af metadata. Der er dog ikke tvivl om, at metadata er et stærkt redskab til at kortlægge adfærd, omgangskreds, vaner mv. samt til brug for profilering. Det dataetiske spørgsmål er, om man kan identificere og informere om brug af metadata, samt om brugeren kan få indsigt og adgang til metadata.

Offentlige platforme

Hvad er dataetisk brug af offentlige platforme?

offentlige platforme, hvor man selv har kontrol over data. Det er vigtigt at have retningslinjer for brug af platformen og moderere den, så individers følsomme persondata så vidt muligt ikke eksponeres.

On-device processing

Hvad er on-device processing?

Frem for at lagre persondata til egen server eller cloud kan man behandle data direkte på brugernes enheder. Det gør Apple med Siri, det gør Wire og Cliqz. Når der er behov for serverbehandling af data, indsamles data anonymt uden at kunne identificeres.

Open source

Hvorfor er open source godt?

Open source betyder, at kildekoden i dine programmer er frit tilgængelig, og at andre kan forbedre, teste, fejlrette og sikre systemet – der er support fra hele verden, og sikkerhedshuller fan-

ges ofte hurtigt. Open source er ikke nødvendigvis gratis men er ofte billigere og gør dig uafhængig af leverandøren. Og så er der tale om gennemsigtighed: Dine brugere – eller tredjeparter – kan selv konstatere, hvad produktet indeholder.

Organisatorisk forandring

Hvad er dataetisk organisatorisk forankring?

Dataetik er ikke et "one man job". Det er en bred tilgang, der dækker alt fra fx produktudvikling, innovation og marketing til strategisk udvikling. Den dataetiske tilgang bør derfor være drevet helt fra topledelsen og være understøttet som en værdi blandt medarbejderne. Det kan gøres på forskellige måder. For eksempel har Apple privacy-eksperter, der er involveret i alle produktudviklingsteams fra starten, og AXA har et board, der bliver fløjet til Paris tre gange om året for at diskutere dataetiske dilemmaer.

Privacy by Design

Hvad er Privacy by Design?

Privacy by Design (PbD) handler om, at standardindstillingen på en tjeneste er privat (private by default), og at den designes og udvikles med privacy som udgangspunkt. De første PbD-principper blev udviklet i 1990'erne af Ann Cavoukian, tidligere direktør for datatilsynet i Canada. Men siden da har de udviklet sig. Man kan også vælge at se PbD som en forretningsfilosofi, dvs. en innovativ tilgang til den digitale forretningsudvikling, hvor privacy er udgangspunktet for alle de forskellige innovative forretningsprocesser, en virksomhed sætter i gang, lige fra design og teknologisk udvikling til human ressource-udvikling, CSR og markedsføring. PbD-principperne bliver på den måde en generel vejledning til at bygge alternativer til den datadrevne public-by-default virksomhedstype.

Profilering

Hvad er profilering?

Profilering er en automatisk behandling af data, der analyserer bestemte personlige forhold vedrørende en person for at lave forudsigelser vedrørende fx dennes arbejdsindsats, økonomiske situation eller helbred. Profilering er en dataetisk udfordring, fordi det er den mest intense form for behandling af persondata, og man skal derfor allerførst afgøre, om der findes et retsgrundlag for profileringen. En dataetisk profilering vil altid være udviklet til fordel for individet, og individet vil have mulighed for at influere og bestemme værdier, regler og input, der bruges i profileringen.

Salg af data

Kan man sælge data til tredjepart på etisk ansvarlig vis?

Det er ulovligt at sælge data til tredjepart, medmindre man har et klart samtykke til det. Et

flertal af danskere tror, at organisationer sælger data videre til tredjepart. Det gør de ikke nødvendigvis, men derfor kan det være vigtigt at forklare dét, man selv tager for givet. Man kan godt sælge data til tredjepart etisk ansvarligt, hvis der er tale om fuldt ud anonymiserede oplysninger baseret på mønstre blandt en gruppe af borgere.

Sårbare

Hvem er særligt sårbare, når det gælder databehandling?

Der indsamles ofte flere data om særligt sårbare mennesker såsom flygtninge, personer med fysisk handicap, psykisk syge, socialt udsatte, arbejdsløse, indsatte osv. Fx indsamler Udbetaling Danmark mange oplysninger om ydelsesmodtagere og deres samlevere og husstand og formodede samlevere og husstand.

Tjene penge

Kan man så ikke tjene penge på databehandling?

Jo, det kan man, hvis mennesket er i centrum. Økonomiske interesser kan aldrig tilsidesætte menneskets grundlæggende rettigheder såsom retten til privatliv, selvbestemmelse og til ikke at blive forskelsbehandlet eller stigmatiseret. Dette kan eksempelvis opnås ved Privacy-by-Design.

Transparens

Hvad er transparens?

Det er ikke nok blot at have en "Transparency Report", der fx handler om, hvor mange gange staten beder om adgang til en virksomheds data med en dommerkendelse. Transparens handler også om egne dataprocesser. Transparens er skridtet videre end basale juridiske krav til samtykke og mulighed for indsigt og indsigelse. Datasystemer og -processer designes til at skabe tillid hos individet samt til at give individet reel

mulighed for at gøre indsigelse mod behandling af data. Det enkelte menneske skal kunne modsætte sig databehandlingen uden at miste rettigheder.

Zero-knowledge

Hvad er zero-knowledge-principper?

Ifølge GDPR må man ikke opbevare data længere, end det er nødvendigt i forhold til formålet. Man kan vælge at gå længere end lovgivningen og slette data, før det er nødvendigt, fordi man ikke ønsker at sidde på risikoen ved at have ansvaret for følsomme data. Man kan fx vælge automatisk helt at slette lokationsdata efter 24 timer, hvis man er et forsikringsselskab, der tracker sine bilforsikringskunder.

Få flere informationer på DataEthics.eu, hvor du kan finde værktøjer og tjenester både til dataetik og digitalt selvforsvar.