A close-up portrait of a man with medium-length, wavy brown hair and light-colored eyes. He is wearing a light blue and white vertically striped button-down shirt. The background is a blurred indoor setting with vertical blinds. The text is overlaid on the lower part of the image.

“Der er simpelthen mange flere mennesker, der arbejder med machine learning i dag end for år tilbage”

Allan Grønlund, specialkonsulent, Datalogisk Institut, Aarhus Universitet



MACHINE LEARNING

går mainstream

Med enorme mængder data og computerkraft i ryggen er machine learning lige nu landet på toppen af hype-cyklussen. Der fremkommer stadig flere anvendelser, som kan få direkte betydning for os alle.

Af Stig Andersen
[prosabladet@prosa.dk]

Machine learning er lige nu blandt de allermest hypede teknologier – ikke blot i snævre branchekredse eller i den akademiske verden. Mainstream-medier kan berette om, at maskinerne snart overtager kontrollen, og at den gamle drøm om kunstig intelligens nu er ved at blive realiseret. Der er dog langt fra tale om ren hype. Amazon, Google, Facebook, Baidu og andre tech-giganter har alle udviklet machine learning-baserede anvendelser inden for deres forretningsområder. Og vi er allesammen med til dagligt at føde mere og mere data ind til algoritmerne. Så snart man laver en rettelser i Google Translate, er man sammen med millioner andre brugere med til gradvis at gøre oversættelsen bedre og bedre – bare for at tage et enkelt eksempel.

Mange open source-værktøjer

Også på andre områder er machine learning-baserede løsninger ved at få direkte indvirkning på vores liv. Medicinsk diagnosticering, scanning og sortering af jobansøgninger, profilering og segmentering ud fra din adfærd på nettet, automatisk sagsbehandling i forsikringsbranchen, osv. osv. Alle anvendelser, der enten allerede er implementeret eller er på vej.

At teknologien er ved at være mainstream, ses også af de efterhånden mange open source-værktøjer og API'er, som man umiddelbart kan tage ned fra hylden og bruge til at udvikle applikationer. Stort set alle de store teknologivirksomheder har lagt machine learning-pakker ud som open source, og de fleste programmeringssprog har deres pakker, eksempelvis scikit-learn til Python og Weka til Java. En simpel søgning på "open source machine learning" vil give adskillige gode lister over de mest populære værktøjer.

Men hvorfor dette boom for en teknologi, hvis grundkoncepter har været kendt i en del år?

De ufatteligt store mængder data, vi har til rådighed i dag – og vel at mærke labeled data, som eksempelvis billeder på nettet – er en vigtig faktor. Machine learning-algoritmerne kan fodres med enorme mængder eksempeldata, som sætter dem i stand til at blive stadig bedre til opgaver, som vi tidligere kun kunne sætte mennesker til. Det kan være alt fra på et millisekund at finde alle matchende ansigter i en billeddatabase til at slå den regerende verdensmester i brætspillet Go.

Dertil kommer den voldsomt øgede maskinkraft, vi har til rådighed i dag, og så en tredje faktor, som Allan Grønlund, specialkonsulent på Datalogisk Institut, Aarhus Universitet, peger på.

– Både i industrien og i den akademiske verden er man blevet meget klogere på det her område de seneste år, og der er simpelthen mange flere mennesker, der arbejder med machine learning i dag end for år tilbage, siger han.

Teknologien ikke ny

Algoritmerne, der ligger i machine learning-værktøjskassen, er ifølge Allan Grønlund ikke specielt nye. På nogle områder er der dog sket en udvikling.

– Der er sket en del ændringer i den måde, hvorpå man opbygger neurale netværk, som er en grundkomponent i den variant af machine learning, vi kalder deep learning, forklarer han.

På de følgende sider kigger vi nærmere på, hvad machine learning er, og ikke mindst ser vi nærmere på nogle af de etiske og juridiske problemstillinger, som en øget anvendelse af machine learning-baserede systemer kan føre med sig.

HVAD ER machine learning?



Kunstig intelligens, neurale netværk, selvlærende systemer. I fem årtier har begreber og teknologier udviklet sig rundt om det, vi i dag taler om som machine learning.

Af Stig Andersen
[prosabladet@prosa.dk]

Allerede i 1959 definerede Arthur Lee Samuel, pioner inden for kunstig intelligens og selvlærende spil, machine learning som "the field of study that gives computers the ability to learn without being explicitly programmed". Det opsummerer grundlæggende begrebet, men der er siden kommet mange og mere detaljerede bud blandt andet fra udviklingshuse og undervisningsinstitutioner.

SAS Institute, som er en af de helt store spillere inden for data analytics, bruger denne formulering på sin hjemmeside: "Machine learning is a method of data analysis that automates analytical model building. Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look."

Denne definition medtager et vigtigt element i machine learning: Læringen opstår ved gennem dataanalyse af et

meget stort antal eksempler at bygge og løbende forbedre en model, der kan bruges af en algoritme til at genkende mønstre og efterfølgende returnere gode forudsigelser og svar på det formulerede problem.

En værktøjskasse

Uddannelsesinstitutioner, der har taget machine learning ind som pensum, har også haft behov for at komme med en kort og præcis definition. Datalogisk Institut, Aarhus Universitet, definerer machine learning således i sin kursusoversigt:

"Machine learning handler om automatiske metoder til at lave præcise forudsigelser eller tage anvendelige beslutninger baseret på tidligere observationer og erfaringer. Machine learning kombinerer matematisk modellering, statistik og sandsynlighedsregning med algoritmik og optimering."

“Machine learning is a method of data analysis that automates analytical model building”

SAS Institute



Denne beskrivelse afslører også, at machine learning i høj grad er tværdisciplinær og kan betragtes som en værktøjskasse, man kan gribe i til udvikling af den specifikke applikation.

Supervised, unsupervised og reinforcement

Machine learning opdeles typisk i nogle underkategorier. Nedenfor ses de tre kategorier, man underviser i på machine learning-kurset på Datalogisk Institut, Aarhus Universitet. Beskrivelserne er baseret på input fra specialkonsulent Allan Grønlund, der underviser på kurset.

Supervised learning

Baseret på eksempeldata, eksempelvis billeddata, der hvert er blevet taget med, hvad det er et billede af, udleder supervised learning-mønstre i eksemplerne, der kan bruges til at udlede den korrekte label for det givne input. Disse mønstre kan funktionen så anvende i forsøg på at genkende data uden label og derved give dem en label. Outputtet af supervised learning-algoritmen er en funktion, der gør netop det.

To af de simpleste supervised learning-algoritmer er en algoritme, som finder den bedste lineære funktion, der passer med eksempeldataene og en funktion, der på et givet input returnerer en label for det punkt i eksempeldataene, der minder mest om inputpunktet.

En anden variant er deep learning, hvor algoritmen finder et godt "dybt neuralt netværk" til at finde mønstre.

Disse dybe neurale netværk kan definere/approximere en bred vifte af funktioner og beskrives typisk som en serie af lag, der hver laver en række lineære kombinationer af outputtet af det tidligere lag og derefter sender resultaterne gennem en ikke-lineær transformation og videre til det næste lag. Så det dybe kommer af, at dataene går igennem mange (ikke-lineære) transformationer for at lave mønstergenkendelsen, inden der tages en beslutning om, hvilken label inputtet skal have. Det er disse dybe netværk, der har været med til at revolutionere billed- og talegenkendelse igennem de seneste år.

Unsupervised learning

Her arbejder algoritmen på unlabeled data, hvor man ønsker at finde mønstre og strukturer, som eksempelvis kan anvendes til at identificere og gruppere objekter, der minder om hinanden. Kan for eksempel anvendes til analyse af kundeadfærd med henblik på at målrette reklameindsatser.

Reinforcement learning

Minder om de to andre, men trækker på spilkonceptet, hvor en agent interagerer med det omkringliggende miljø, tager en bestemt aktion, hvorved der sker en ændring i miljøet. Algoritmen "lærer ved at spille spillet" og optimeres dermed til at opnå noget bestemt. Anvendes eksempelvis i kontrolsystemer og tenderer noget, man kunne kalde kunstig intelligens.

AUTOMATI

er på vej i kommunerne

Kommunerne er i gang med at automatisere administrative processer. Machine learning-baserede systemer til sagsbehandling ligger dog et godt stykke ude i fremtiden.

Af Stig Andersen
[prosabladet@prosa.dk]

Hvis man er bekymret for, at afgørelser i den offentlige sagsbehandling snart bliver spyttet ud af machine learning-baserede sagsbehandlingsløsninger, er man ude i meget god tid ifølge nogle af de folk, der i kommunalt regi arbejder med automatisering.

– Vi er startet med nogle meget lavthængende frugter, hvor vi benytter nogle af de almindelige machine learning-algoritmer til at øge præcisionen af vores analyser, eksempelvis til at automatisk tjek af, om en given fil er af en bestemt dokumenttype, fortæller Jonas Kristian von Lüttichau, chefkonsulent it-arkitektur, Københavns Kommune.

Kontrol og risikovurdering

Kommunen er med i et ph.d.-projekt på Datalogisk Institut, Københavns Universitet, hvor en ph.d.-studerende skal udvikle nye og hurtigere machine learning-algoritmer. Planen er, at man undervejs i projektet skal kunne anvende nogle af algoritmerne på afgrænsede forretningsområder i kommunen. Første use-case ligger inden for sygedagpenge og handler om at gennemløbe de ansøgninger om refusion hos staten, som en kommune skal indsende. Man forestiller sig, at output er en risikovurdering af, hvorvidt ansøgningen er fyldestgørende nok til at gå igennem.

– Vi har med vilje valgt at starte med noget, der er en helt intern kontrolfunktion og altså ikke berører borgeren direkte, og som kun trækker på ét datasæt. På den måde undgår vi i det konkrete projektforsøg at skulle tage stilling til nogle mere overordnede principielle eller juridiske problemstillinger, fortæller Jonas Kristian von Lüttichau.

Han er helt med på, at machine learning anvendt inden for offentlig sagsbehandling kan give anledning til

nogle overvejelser i forhold til privacy, transparens og lignende, men han mener, at det ligger noget længere ude i fremtiden.

– Vi er slet ikke der, hvor robotterne kommer, og hvor kunstig intelligens er ved at overtage det hele. Lige nu er vi bare på et meget grundlæggende plan ved at finde ud af, hvilke muligheder vi har, og på hvilke forretningsområder det kunne give mening at anvende disse teknologier, siger han og fortsætter:

– Jeg synes, det ville være en skam at undlade at undersøge mulighederne for at forbedre vores arbejdsgange og skabe mere kvalitet også i det borgervendte arbejde, fordi der engang kunne opstå nogle principielle problemstillinger.

Han fortæller, at når han er rundt i organisationen for at



ISERINGEN

“Jeg synes, det ville være en skam at undlade at undersøge mulighederne for at forbedre vores arbejdsgange og skabe mere kvalitet også i det borgervendte arbejde, fordi der engang kunne opstå nogle principielle problemstillinger”

Jonas Kristian von Lüttichau, Chefkonsulent IT Arkitektur, Københavns Kommune

fortælle om projektet og afsøge mulige anvendelsesområder, bliver der stort set altid spurgt til eventuelle problemstillinger i forhold til privacy og datasikkerhed. Det er også problemstillinger, som man er meget bevidst om i det Big Data-team, han er leder af.

– Udviklerne i teamet er meget bevidste om, at deres arbejde skal foregå inden for rammer af, hvad der er etisk og juridisk i orden, siger han.

Mest dataindsamling

I Helsingør Kommune er man også i gang med at afsøge mulighederne i de nye teknologier. I øjeblikket er man ved at identificere områder, hvor det kunne give mening at automatisere arbejdsgangene ved hjælp af softwarerobotter, primært til dataindsamling og generering af rapporter.

– Lige nu er vi i en fase, hvor vi ser på mulighederne for at sætte strøm til nogle gentagne manuelle processer, så vores sagsbehandlere kan bruge deres tid der, hvor det giver mere mening, fortæller Peter Christensen, leder af digitalisering i Helsingør Kommune, som arbejder med nye digitale muligheder sammen med seks andre kommuner i Nordsjællands digitaliseringssamarbejde.

Lige nu presser de store principielle problemstillinger sig ikke på.

– Det, vi gør lige nu, handler om dataindsamling, hvilket

vi jo har gjort gennem længere tid, eksempelvis i vores ledelsessystemer eller for den sags skyld manuelt. Den dag, vi måtte begynde at anvende systemerne til at træffe afgørelser, skal vi dog helt sikkert gøre os en række vigtige overvejelser, eksempelvis i forhold til transparens, siger Peter Christensen.

Han forudser, at hvis løsninger bliver anvendt i forbindelse med sagsbehandlingen, bliver det formentlig som en støtte for sagsbehandleren og ikke en automatisk genereret afgørelse, der bliver sendt direkte ud til borgeren.

Alle forudsætninger i Danmark

Morten Lindblad er direktør for Dataproces, som er et af de firmaer herhjemme, der er meget aktive på det offentlige marked for automatisering af administrative processer, og som udvikler og driver Big Data-analyseløsninger, hvor machine learning helt naturligt indgår.

– I øjeblikket er der ikke implementeret offentlige projekter, hvor der er tale om egentlig machine learning-baseret sagsbehandling. Softwarerobotterne, som man kunne kalde første generation på det område, er til gengæld i stor vækst, fortæller han og fortsætter:

– Vi har alle forudsætninger i Danmark for at kunne udnytte mulighederne i machine learning, da vi har en meget høj grad af digitalisering. Men det ligger stadig et godt stykke ude i fremtiden.

Moral behandlet som

MATHEMATIK

De grundlæggende moralske og juridiske spørgsmål forsvinder ikke, bare fordi man pudser machine learning-algoritmer på dem. Automatiseringen åbner til gengæld for nye.

Af Stig Andersen
[prosabladet@prosa.dk]

Vi har i mange år været vant til at automatisere processer, der ikke krævede det, vi opfatter som menneskelig viden, erfaring og intelligens. Det er ved at ændre sig i disse år, hvor computerdrevne systemer i stigende grad bliver taget i anvendelse på områder, der tidligere var forbeholdt mennesker. Risikovurdering i forsikringsbranchen, kreditvurdering i finanssektoren og medicinsk diagnosticering er blot nogle af de områder, hvor mennesket helt eller delvist er ved at blive erstattet af avancerede machine learning-baserede systemer, der bliver bedre, jo mere de bliver brugt.

Machine learning bliver ofte koblet til den efterhånden gamle vision om kunstig intelligens. Sammen med billedet af "maskiner, der overtager kontrollen", har det rejst en række mere principielle spørgsmål, som har fyldt meget i den bredere debat om teknologien. Direktøren på Oxford University-instituttet Future of Humanity Institute, professor Nick Bostrom, var i et interview med The Guardian ude med noget nær dommedagsprofetier om menneskets undergang, når maskinerne bliver klogere end os og på evolutiønær vis vil overtage verdensherredømmet fra mennesket.

Meget mindre dramatisk, men lidt i samme spor er Cathy O'Neil, matematiker på Harvard, som i sin bog "Weapons of Math Destruction" formulerer bekymringen for de ødelæggende konsekvenser, hvis man i for høj grad stoler på, at matematiske modeller kan anvendes til at tage beslutninger med store konsekvenser for det enkelte menneske.

Trussel mod transparensen

Denne debat er også taget op herhjemme, blandt andre af professor på IT-Universitetet Rasmus Pagh. Han peger i en artikel på itu.dk på, at der er en risiko for, at det kan gå ud over transparensen i beslutningsprocesserne i både offentligt og privat regi. Risikoen opstår, hvis vigtige afgørelser i stadig højere grad bliver taget på basis af så mange parametre og af så avancerede algoritmer, at kun et fåtal efterhånden vil kunne gennemskue dem.

Samme problemstilling peger Flemming Bjerke, bestyrelsesmedlem i IT-Politisk forening, på.

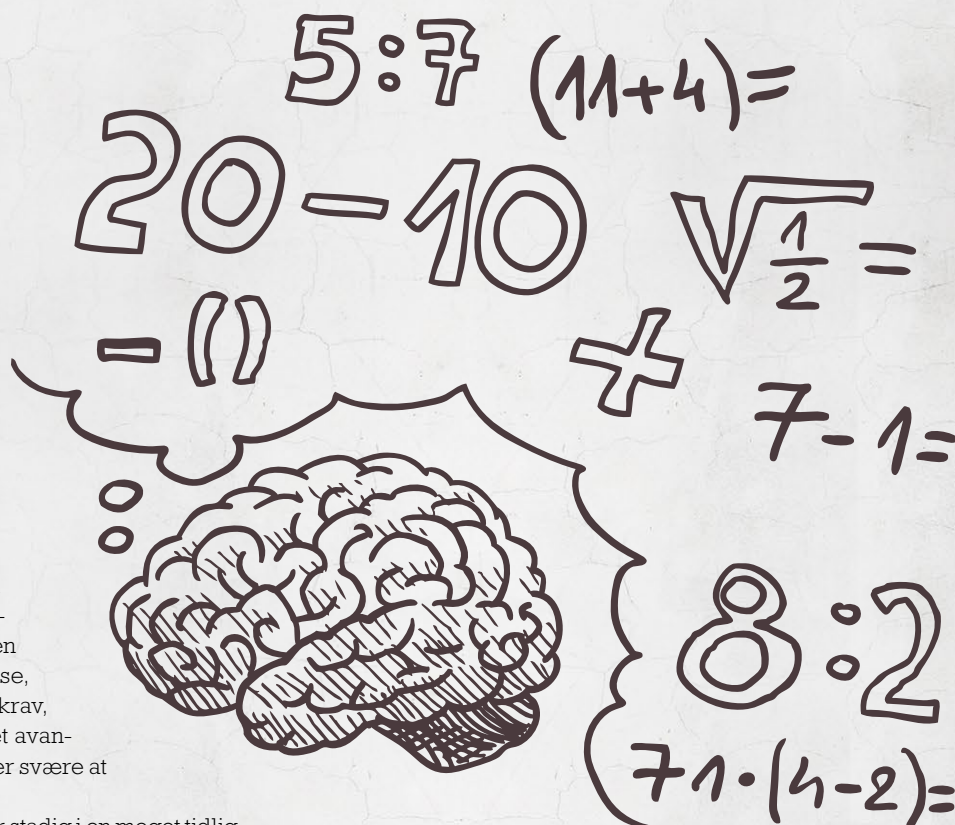
– Det er i forvejen meget svært for den enkelte at gennemskue, dels hvilke data der bliver indsamlet om en, dels

hvad de bliver brugt til. Det problem vil bare blive større, jo mere vi lader machine learning-baserede systemer træffe afgørelser, da de indsamlede data dermed bliver brugt til stadig flere formål, siger han.

Ret til forklaring

Inden for offentlig forvaltning er der en skærpet problemstilling i forhold til transparens. Som borger har man et retskrav på at få en forståelig forklaring på en afgørelse, og hvordan kan man honorere det krav, hvis afgørelserne er taget af meget avancerede systemer, hvis præmisser er svære at gennemskue?

Automatiseret sagsbehandling er stadig i en meget tidlig fase i det offentlige, men Niels Fenger, professor i forvaltningsret ved Københavns Universitet, vurderer umiddel-



– Helt grundlæggende forudsætter vores lovgivning jo en eller anden form for menneskelige egenskaber og beslutningsprocesser. Det er jo den målestok, vi har været vant til at holde en given handling op imod. Og hvordan foretager man den vurdering, hvis det er en computer, der har taget beslutninger, som ligger længere og længere væk fra noget, et menneske har truffet beslutning om?, spørger han.

Henrik Udsen vurderer, at de nuværende regler for ansvarsplacering er tilstrækkelige, men vil ikke afvise, at det kan blive nødvendigt at justere dem på et tidspunkt.

– Vi ved jo reelt ikke, hvor teknologien ender. Hvor ligger ansvaret, hvis computere begynder at tage beslutninger, som vi slet ikke kunne forudse, da vi bragte dem i anvendelse? Her bliver det i virkeligheden mindst lige så vigtigt at have tilstrækkelige kontrolprocedurer i forhold til, hvordan en given løsning udvikler sig, så man har mulighed for at justere på algoritmerne, siger han.

Ret til "menneskelig indgriben"

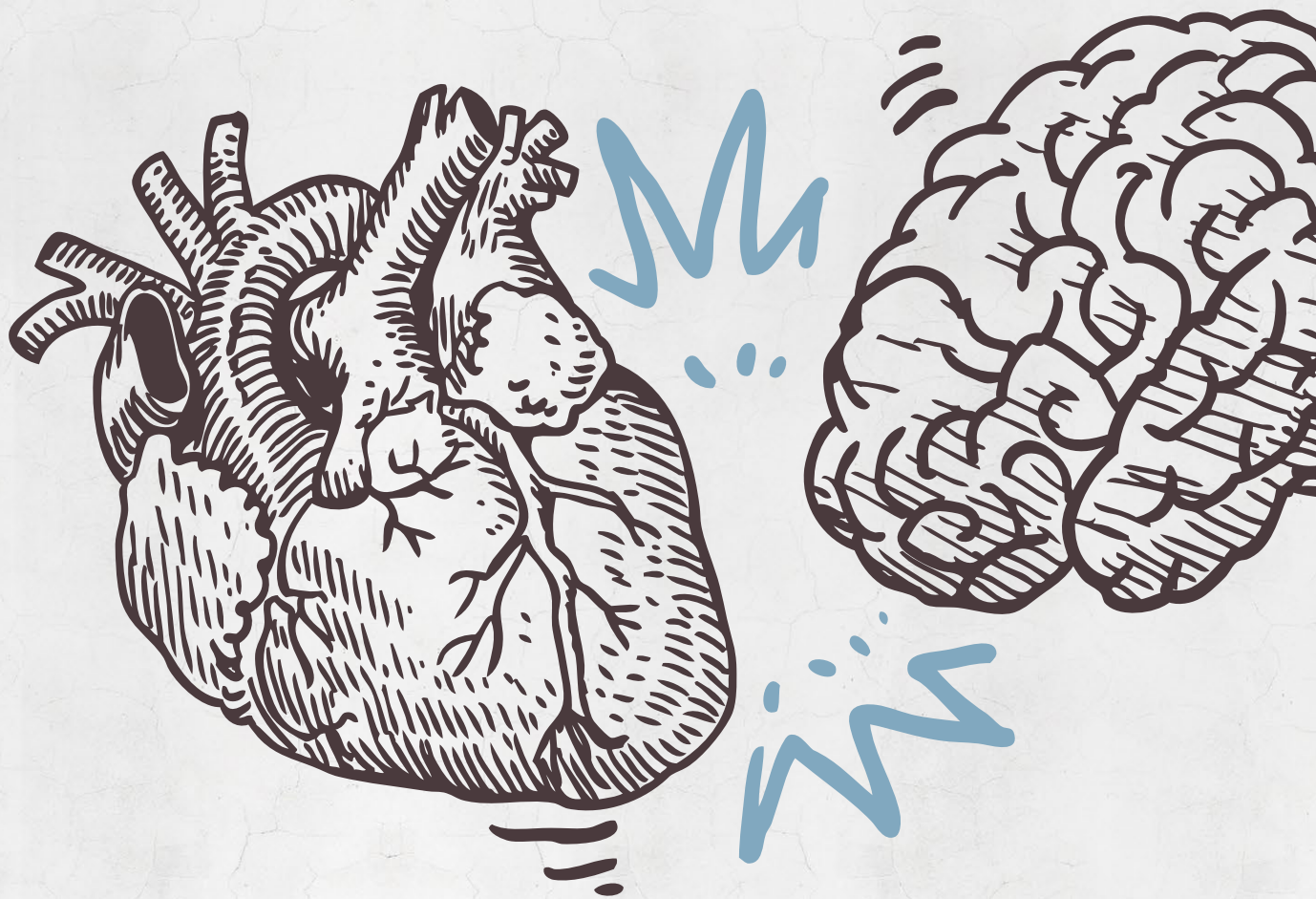
Den nye EU-persondataforordning, som træder i kraft 25. maj 2018, forholder sig til spørgsmålet om afgørelser, som baserer sig på en automatisk sagsbehandling. I artikel 71 hedder det, at man "bør have ret til ikke at blive gjort til genstand for en afgørelse, der kan omfatte en foranstaltning, som evaluerer personlige forhold vedrørende vedkommende, og som alene bygger på automatisk behandling, og som har retsvirkning, eller som på tilsvarende vis betydeligt påvirker den pågældende, såsom et automatisk afslag på en onlineansøgning om kredit >>

bart, at den eksisterende begrundelsespligt er tilstrækkelig også i en situation, hvor afgørelser bliver truffet af selv-lærende systemer.

– Forvaltningen er vant til at skulle begrunde afgørelser, og hvis man ikke er i stand til at forklare en afgørelse, har man under alle omstændigheder et problem. Rent juridisk vil det i sidste ende jo stå og falde med, om afgørelsen er lovlig eller ej - ikke om det er en maskine eller et menneske, der har truffet den, siger han.

Hvem har ansvaret?

Udbredelsen af automatiserede løsninger, der "selv tager beslutninger", kan potentielt også udfordre hele spørgsmålet om ansvarsplacering, mener Henrik Udsen, professor i it-ret ved Københavns Universitet.



eller e-rekrutteringsprocedurer uden nogen menneskelig indgriben". Til gengæld ligger der ikke noget krav om, at borgeren skal oplyses om, at en afgørelse er truffet i en automatiseret proces.

Risiko for bias

En risiko ved machine learning-baserede systemer, der ofte bliver peget på, er risikoen for diskrimination, når man anvender systemerne til profilering på basis af personoplysninger. Profilering er meget udbredt som middel til at målrette reklamekampagner, men bliver anvendt til en lang række andre formål. Et eksempel er en løsning fra det amerikanske firma Northpointe, som bliver anvendt i mange offentlige institutioner USA til at forudsige, hvem der på et tidspunkt bliver kriminelle – en løsning, som har en kraftig bias mod sorte.

Problemet er, at det ikke er muligt at få indsigt i, hvilke præmisser og hvilken model som ligger til grund for de algoritmer, der bliver anvendt. Så ud over den diskrimination, der kunne være tale om, så illustrerer eksemplet en generel problemstilling i forhold til transparens: Mange algoritmer, der bliver anvendt i det offentlige, er ejet af private virksomheder, som ikke ønsker at offentliggøre de bagvedliggende modeller. Og selvom de gjorde, er der ifølge Ole Tange, it-politisk rådgiver i PROSA, stadig et problem:

– Selvom man gav borgerne adgang til at se, hvordan "maskinen træffer beslutninger", så kan dette være så komplekst, at man som menneske kan have meget svært ved at gennemskue det.

Handler om moral

Der bliver skrevet og sagt meget om, hvad machine learning-systemer efterhånden vil være i stand til. For Flemming Bjerke, IT-Politisk forening, er der et andet spørgsmål, som er meget vigtigere.

– Der er ingen grund til at tro, at denne type systemer ikke efterhånden vil kunne tage sig af opgaver, som vi i dag tror, at kun mennesker kan klare. Så det vigtige spørgsmål er, om vi overhovedet ønsker at sætte et system til at træffe vigtige afgørelser for mennesker. I praksis er det et problem for retssamfundet, siger han og fortsætter:

– Interesser, partsindlæg, tolkninger, vinklinger og uforudsigelighed findes jo stadig, selvom man bruger kunstig intelligens - uanset at den kunstige intelligens måske vil undertrykke det.

Cathy O'Neil gør også op med forestillingen om, at den bagvedliggende matematik i machine learning-systemer på nogen måde skulle være garant for neutralitet og objektivitet.

– I sidste ende handler det om moral, værdier og prioriteter. Grundlæggende er det et spørgsmål om moral, der



“I sidste ende handler det om moral, værdier og prioriteter. Grundlæggende er det et spørgsmål om moral, der bare bliver behandlet, som om det var matematik”

Cathy O'Neil, matematiker, Harvard Universitet

bare bliver behandlet, som om det var matematik. Hvorfor? Fordi folk stoler på matematik. Hvis man kan pakke noget ind i matematik, vil folk holde op med at stille spørgsmål, siger hun i et interview i Politiken.

Altid en fall-back

Ole Tange ser også en risiko for, at den svagere part, typisk den enkelte borger, kan risikere at blive klemt.

– Machine learning vil kunne give fornuft i rigtig mange forskellige situationer og kan formentlig være en god støtte for sagsbehandling. Men der skal altid være en fall-back i de sager, som af den ene eller anden grund falder uden for det, der kan rummes af den automatiserede løsning. Og det skal være en reel mulighed: Det må ikke bare være en teoretisk mulighed, som i praksis er uden for svage borgeres formåen, siger han og fortsætter:

– Til de uproblematisk sager kan man nok godt lave en 100 procent automatiseret behandling, men til mere komplekse sager vil jeg anbefale, at automatiseret sagsbehandling bliver en støtte for en sagsbehandling, hvor der også er et menneske inde over for at lave en vurdering. Maskinens rolle vil derfor blive reduceret til at finde de fakta, som underbygger en given beslutning, men det er mennesket, som baseret på disse fakta træffer beslutningen.

Ensartethed og højere kvalitet er de to punkter, der typisk nævnes på plussiden i forhold til automatiserede sagsbehandlingsløsninger.

– Hvis systemet er lavet ordentligt, og man anvender det på noget, der er simpelt nok, kan det være en kæmpe fordel, da det kan sikre en ensartethed og kvalitet i sagsbehandlingen, måske ikke mindst i form af understøttende systemer, der gør sagsbehandleren opmærksom på vigtige forhold, siger professor Niels Fenger.

Flemming Bjerke mener også, at det vil give fin mening at sætte et system til at håndtere de mere rutineprægede afgørelser, og ensartethed i sagsbehandlingen ser han også som en fordel.

– Det kan godt være, at der er en risiko for systemmæssig bias mod en bestemt befolkningsgruppe. Modsat kan det jo også være et værn imod, at en sag uheldigvis bliver behandlet af en medarbejder, der har fordomme over for bestemte grupper, siger han.

Til gengæld giver han ikke meget for forestillingen om, at man vil nøjes med at lade systemerne være en støtte for en menneskelig sagsbehandler.

– For det første vil det ikke give mening at lade afgørelserne mellemlande på et sagsbehandlerbord, hvis systemerne rent faktisk er gode nok til at håndtere sagerne. For det andet viser erfaringerne, at hvis der kan laves besparelser gennem automatisering, så vil det blive gjort, siger Flemming Bjerke.