

Ayo Næsborg-Andersen, Jøren Ullits og Ole Hammerslev

Beslutningsunderstøttende algoritmer i det offentlige: påvirkningen af sagsbehandlerens skøn og begrundelse

Sagsbehandlerne i forvaltningen skal træffe afgørelser på et oplyst og objektivt grundlag og være i stand til at begrunde afgørelserne. De assisteres dog i stigende grad af beslutningsunderstøttende algoritmer, der er skabt til at generere output fremfor at tilvejebringe en forklaring. Denne udvikling kan effektivisere og forbedre arbejdsgange, men kan også udfordre forvaltningsretlige krav til skøn og begrundelse, der hviler på grundlæggende retssikkerhedsmæssige principper om blandt andet en tilstrækkelig sagsoplysning og forudberegnelighed. Ved at bruge ASTA-rapportens data om sagsbehandleres oplevelser af og praksis med brugen af algoritmer i risikovurderingen af nyledige illustreres, hvordan *automation bias* og *black box*-paradokset kan påvirke sagsbehandlerens skøn og begrundelse. Så på trods af en måske mere effektiv håndtering af større datamængder via algoritmer kan det udfordre nogle af retsstatens grundlæggende principper.

Nøgleord: fagprofessionelt skøn, vurderingsskøn, begrundelse, algoritmer, sagsbehandling, *automation bias*, *black box*-paradokset

En række danske jobcentre har forsøgt at risikoscreene for potentiel langtidsledighed ved at benytte en række variabler via programmet ASTA. På baggrund af ASTAs variabler og den vægtning af variablerne, som udviklerne har fundet optimal for den bagvedliggende model, inddeler ASTA de nyledige i farvekodeerne rød, gul eller grøn alt efter risikoen for langtidsledighed. Inddelingen har betydning for, hvor mange resurser jobcentret skal allokere til den pågældende nyledige. Det er interessant, at flere sagsbehandlere fra jobcentrene ikke altid kan se en logisk sammenhæng mellem ASTAs risikoscreening og den konkrete sag og dermed ikke kan forklare, hvorfor borgerne bliver kategoriseret, som de gør. Ydermere er sagsbehandlerne usikre på, om ASTA er mere præcis til at identificere risiko, end de selv er, og de er i tvivl om, hvorvidt de kan stole på korrektheden af ASTAs tolkning af den nyledige (Flügge et al., 2022).

Et af formålene med en mere digital forvaltning er, at algoritmer kan hjælpe forvaltningen med at sikre effektive, lovlige og retfærdige afgørelser eller sågar selv træffe afgørelser, men eksemplet illustrerer, at det kan være svært, endog helt umuligt, at gennemskue, hvilke mellemregninger der ligger til grund for

algoritmens vurdering. Denne udfordring er bedre kendt som *black box*-paradokset (Pasquale, 2015). Det betyder, at den måde, AI profilerer borgere på, kan være ganske ugennemsigtig for den enkelte sagsbehandler. Selvom ASTA alene er et beslutningsunderstøttende værktøj, når sagsbehandlere skal foretage deres samlede skøn, er det interessant, at sagsbehandlerne ikke nødvendigvis kan gennemskue ASTAs kategoriseringer og hermed kan have vanskeligt ved at forstå den systemgenererede vurdering, der i eksemplet med ASTA forudsættes styrende for sagens forløb. De kan ligeledes have vanskeligt ved at applicere og udfolde denne vurdering i en afgørelsessammenhæng som påkrævet efter forvaltningslovens begrundelseskrav (Flügge et al., 2022).

At der ikke er klarhed over det datagrundlag, som sagsbehandleren baserer sit skøn på, strider mod hele forestillingen om en forvaltning, der både skal træffe lovmæssige afgørelser på et oplyst grundlag og være i stand til at sammenfatte dette grundlag til en begrundelse. Gennemsigtighed om, hvad myndigheden har tillagt vægt, tilvejebringes netop gennem begrundelsen, der derved bliver afgørende for borgerens vurdering af, om sagen skal ankes, og tillige for ankeinstansens eventuelle efterprøvelse af førsteinstansens behandling af sagen. Udfordringen med beslutningsunderstøttende algoritmer som ASTA er i denne sammenhæng, at sagsbehandlerne har vanskeligt ved at efterprøve den vurdering, et AI-system som ASTA genererer, og som sagsbehandleren forudsættes at inddrage i vurderingen af sagens videre forløb. Ligeledes bliver det derved vanskeligt at forklare overfor borgeren, hvorfor man har valgt at følge en vurdering, som man måske ikke til fulde forstår baggrunden for. Og denne forståelse er påkrævet, hvis forvaltningen ønsker at bruge vurderingen i en afgørelse, hvor der gælder et krav om begrundelse.

Derfor er det relevant at undersøge forholdet mellem algoritmer og sagsbehandlerens skøn på den ene side og sagsbehandlerens begrundelse på den anden side. For at blive klogere på dette forhold spørger vi i denne artikel: *Hvordan påvirker beslutningsunderstøttende algoritmer sagsbehandlerens skøn og begrundelse af afgørelsen?*

Ved ”skøn” forstås i artiklen et såkaldt vurderingsskøn, hvilket dækker over de situationer, hvor sagsbehandleren er nødt til at vurdere validiteten af en given information. Begrebet forklares nærmere senere.

I det følgende afsnit diskuteres den juridiske litteratur, som denne artikel skriver sig ind i. Selvom der eksisterer litteratur om hhv. skønnet og den digitale forvaltning, ses der ikke noget, der kombinerer de to problemfelter, endsi- gte fokuserer på algoritmernes betydning for sagsbehandlerens skøn. Artiklen har derfor som formål at bidrage til at lukke dette hul. Efter en præsentation af artiklens metoder redegøres for de retlige rammer for skøn og begrundelse i

forvaltningen og i forlængelse deraf for algoritmer og risikofaktorer ved algoritmisk understøttede afgørelser. I den efterfølgende analyse undersøger vi med udgangspunkt i ASTA-casen, hvordan beslutningsunderstøttende algoritmer kan påvirke sagsbehandleres skøn og begrundelse. Endelig diskuteres artiklens fund.

Forskningsmæssig relevans og afgrænsning

Der findes en del juridisk litteratur om skønnets rolle i forvaltningen (se fx Rønsholdt, 2018: 151; Bønsing, 2018: 316ff; Fenger, 2018: 316-320, 367-370; Mørup et al., 2022: 213-277). Enkelte behandler digital forvaltning (især Fenger, 2018), men ingen af værkerne indeholder væsentlige overvejelser om digitaliseringens betydning for forvaltningens skøn og begrundelse af afgørelser. Selv den mest dybdegående nyere behandling af forvaltningens skøn, en ph.d.-afhandling om skøn under regel, nævner kun digitalisering få gange uden at diskutere, hvordan digitaliseringen påvirker sagsbehandlerens skøn (Stausholm, 2015: 141, 221, 266).

Juridisk forskning i digital forvaltning i Danmark har primært fokuseret på værdibaseret design af den digitale forvaltning (Motzfeldt, 2015), tilsyn med sagsbehandlede it-løsninger og delegation i forbindelse med udvikling af it-løsninger (Motzfeldt, 2017; Motzfeldt og Næsborg-Andersen, 2018; Næsborg-Andersen og Motzfeldt, 2019). Derudover er der skrevet enkelte artikler og bøger om digital forvaltning (Loiborg og Søgaard, 2018; Motzfeldt og Abkenar, 2019; Motzfeldt, Ullits og Kjellerup, 2020; Motzfeldt et al., under udgivelse). Fællesnævneren for denne forskning er, at den først og fremmest fokuserer på lovligheden af forskellige løsninger (se fx også Wisborg, 2022). ASTA er konkret omtalt i to artikler (Loiborg 2020; Motzfeldt, 2020). Begge artikler er dog skrevet før ASTA-rapporten (Flügge et al., 2022) og behandler generelle problemstillinger ved forvaltningens brug af algoritmer. Den juridiske forskning i digital forvaltning har overordnet været mindre optaget af interaktionen mellem menneske og algoritme, herunder hvordan brugen af en beslutningsunderstøttende algoritme påvirker sagsbehandlerens skøn samt udfordringerne forbundet med at bruge algoritmisk tilvejebragte oplysninger i en afgørelse, når sagsbehandleren måske ikke til fulde kan gennemskue konteksten. Litteraturen er således ganske sparsom vedrørende de udfordringer, algoritmer kan give i forhold til myndighedernes begrundelsespligt (se dog Olsen et al., 2019, der argumenterer for, at man i forhold til begrundelsespligten ikke bør stille strengere krav til algoritmen end til den menneskelige sagsbehandler). Internationalt har der heller ikke været udbredt fokus på algoritmers betydning for sagsbehandlerens skøn og begrundelse. Den internationale litteratur består primært af

undersøgelser af, hvordan sagsbehandlere bør bruge beslutningsunderstøttende algoritmer (se fx Meilvang og Dahler, 2022). Denne artikel bidrager derfor til et underudviklet område, idet den undersøger algoritmers påvirkning af sagsbehandlernes skøn og den efterfølgende begrundelse i sager med et algoritmisk underbygget datagrundlag. Ydermere kombinerer artiklen utraditionelt eksPLICIT den juridiske forskning i skønnets og begrundelsens rammer med empirisk viden om sagsbehandlernes oplevelse af beslutningsunderstøttende algoritmers påvirkning heraf.

Implicitt heri afgrænser vi os således til alene at undersøge beslutningsunderstøttende algoritmer (modsat fuldautomatiserede algoritmer uden menneskelig involvering), og hermed fokuserer vi på, hvordan denne type algoritmer påvirker skønnet. Sagsbehandlere er dog ikke frit svævende individer, de indgår altid som en del af en større organisation, der påvirker deres arbejdsmønstre, som kan påvirke skønnet, ligesom en række andre faktorer kan have indflydelse på de betingelser, sagsbehandlere har for at udøve skøn (se fx Nielsen og Hammerslev, 2022). Disse faktorer undersøges ikke, da artiklen tager udgangspunkt i sagsbehandlernes oplevelser og forståelser som beskrevet i ASTA-rapporten.

Metode

For at kunne kvalificere en analyse af algoritmers påvirkning af skøn og begrundelse er det nødvendigt først at få afdækket, hvad de to begreber juridisk dækker over. Denne afdækning vil blive foretaget ud fra en diskussion af den forvaltningsretlige litteratur. Vi har valgt at tage udgangspunkt i alment anerkendte forvaltningsretlige værker, da de er repræsentative for retsvidenskabens forståelser. Da retten er rammesættende for sagsbehandlernes skøn og begrundelser, vil vi benytte den retlige forståelse til at analysere AI's påvirkning – vel vidende at ret i praksis ikke er det samme som ret i bøgerne (se således Pound, 1910) – og sammenholde denne med empirisk viden om sagsbehandlernes oplevelser af og praksis med brugen af algoritmer i risikovurderingen af nyledige. Den empiriske viden baseres på rapporten *Er du grøn? Algoritmer til beslutningsstøtte i det offentlige* af Flügge et al. fra 2022 (ASTA-rapporten), som er refereret i indledningen. ASTA-rapporten undersøger sagsbehandlernes oplevelser af og praksis med brugen af algoritmer i risikovurderingen af forsikrede nyledige. Rapporten er baseret på observationer og 16 interview med fire sagsbehandlere enten som del af observationerne (ti timer) af sagsbehandlernes arbejde i jobcentret i Lyngby-Taarbæk Kommune eller som video-/telefoninterviews. Rapportens forskningsdesign fokuserede på at illustrere sagsbehandlernes daglige arbejde med at risikovurdere nyledige dagpengemodtagere. Fokus var på at forstå de faktorer, som sagsbehandlerne oplever som centrale for værdien ved

ASTA i deres daglige arbejde. Rapporten kan hermed danne en forståelse for sagsbehandlerens inddragelse af ASTA som beslutningsunderstøttende værktøj i sager, hvor skøn er væsentligt for sagsbehandlingen. Beslutninger i jobcentre træffes på baggrund af skøn af individuelle forhold. Samtidig er beskæftigelsesområdet lovmæssigt komplekst, hvor en række lovgivninger har relevans, fx forvaltningsloven, serviceloven, lov om aktiv beskæftigelsesindsats, regler om dagpenge og om kontanthjælp. I det undersøgte jobcenter kan sagsbehandlerne inddrage ASTA som supplement til deres skøn. Rapportens analyse er foretaget med udgangspunkt i ”interpretive metoder”, hvor blandt andet feltnoter og citater er brugt til at underbygge forskellige emner (Flügge et al., 2022: 5).

Vi benytter med andre ord rapporten som en form for *exemplifying case* (Bryman, 2016) for at kunne få kvalificeret empirisk viden til at diskutere sagsbehandlerens skøn og begrundelse. På trods af at rapporten bygger på et begrænset antal interview og observationer, reflekterer den alligevel reelle oplevelser med algoritmer, som kan relateres til sagsbehandlerens skøn. På den vis får vi viden om, hvordan algoritmerne opleves af sagsbehandlere ift. deres praksis, og dette relaterer vi således til den retlige ramme om skøn og begrundelse. Med denne tilgang forsøger artiklen at sætte både juraen og viden om sagsbehandlerens praksis i spil for at få en forståelse for algoritmernes påvirkning af sagsbehandlerens skøn.

ASTA som maskinlæringsmodel

ASTA er en maskinlæringsmodel, som på undersøgelsestidspunktet inddrog 50 variabler (Flügge et al., 2022: 7). Disse bliver holdt op imod historiske sammenlignelige data, hvorefter modellen beregner en risikoscore og derefter giver den individuelle ledige en farve som enten grøn, gul eller rød, alt efter hvor stor risiko den vurderer, der er for langtidsledighed. Variablerne er vidt forskellige og omfatter fx helbredsudfordringer, alder, og hvorvidt den ledige har sagt ja til at modtage sms fra jobcentret.

ASTA er en beslutningsunderstøttende algoritme designet til at bidrage til den faglige vurdering af, hvorvidt der er ”risiko for langtidsledighed”. ”Risiko for langtidsledighed” er således her et skønsmæssigt udtryk, hvis udfyldning baseres på de nylediges konkrete forhold og indtastninger med det formål at etablere en score for den pågældendes risiko for langtidsledighed. Det er med andre ord et system, der skal assistere sagsbehandleren ved at sætte en række variabler på formel og derved effektivisere sagsbehandlerens skønsmæssige vurdering.

Skøn og begrundelse

Skøn

Der findes forskellige typer af skøn. Et skøn kan relatere sig til en uklarhed i en juridisk regel, hvor denne uklarhed søges udfyldt/fastlagt (forvaltningsretligt skøn), men skønnet kan også relatere sig til et ikkenormativt indhold, som fx en oplysning eller en vurdering. Et sådant skøn omtales i litteraturen som ”faktumskøn”, ”værdiskøn” eller ”vurderingsskøn” (se fx Mørup et al., 2022; Bønning, 2018: 320). Vi har valgt at bruge begrebet vurderingsskøn i denne artikel, idet der er tale om en vurdering af, hvorvidt og i hvilket omfang algoritmens resultat kan benyttes (se også Fenger (2018: 319), som omtaler vurderingsskøn som ”en opgave, der går ud på at fastlægge et forhold af faktisk karakter ved et skøn, der kommer så tæt på, hvad der må antages at være korrekt, som muligt”).

Et vurderingsskøn anvendes, når sagsbehandleren som et led i sagsoplysningen står tilbage med ufuldstændige eller måske upålidelige oplysninger. Fx hvis en borger har givet myndigheden oplysninger af betydning for sagen under omstændigheder, der rejser tvivl om oplysningens validitet, uden at myndigheden har mulighed for på anden måde at tilvejebringe klarhed herom, fx ved partshøring af andre parter mv. I denne situation er myndigheden nødt til at foretage en selvstændig vurdering af oplysningens validitet. Denne skønstype kan give anledning til udfordringer i sager som ASTA, hvor systemet tilvejebringer en vurdering, som sagsbehandleren efterfølgende kan være nødt til at forholde sig til og herved foretage et vurderingsskøn over. At der skal foretages et skøn, betyder dog ikke, at det står sagsbehandleren helt frit at træffe en beslutning. Vigtigst i denne forbindelse er et af retsstatens væsentligste principper, nemlig magtfordrejningsprincippet som blandt andet forbyder inddragelse af usaglige hensyn i skønnet (jf. Mørup et al., 2022: 244, 471). De lovlige hensyn kan være generelle på tværs af forskellige sagsområder, som fx hensyn til borgerens retsbeskyttelse, men de kan også være specifikke for en bestemt lovgivning eller en bestemt myndighed. Hvis en sagsbehandler inddrager hensyn, der ikke er direkte relevante for den lovgivning, en afgørelse er truffet på baggrund af, vil der være tale om magtfordrejning, med den konsekvens at afgørelsen anses for ugyldig (Mørup et al., 2022: 327).

Forvaltningen skal også overholde lighedsprincippet. Man kan anskue lighedsprincippet som en del af magtfordrejningslæren, som et forbud mod usaglig forskelsbehandling (Mørup et al., 2022: 249-250). Lighedsprincippet indebærer, at forvaltningen skal behandle ensartede sager på samme måde. Er der eksempelvis etableret en fast praksis for, hvordan bestemte typer af sager afgøres, følger det af lighedsprincippet, at skønnet begrænses af denne praksis, med dets krav om lige behandling af det lige og ulige. Der er sjældent to sa-

ger, som er helt identiske, men er omstændighederne og forholdene sammenlignelige, så fastslår lighedsprincippet krav om ensartet behandling. Tilsvarende gælder for myndighedernes brug af beslutningsunderstøttende algoritmer, idet de forvaltningsretlige regler og principper ideelt er teknologineutrale. Hvis den beslutningsunderstøttende algoritme eksempelvis kommer med en anbefaling til sagens videre forløb, der er uforenelig med hidtidig praksis, så har sagsbehandleren ikke alene ret, men også pligt til at undlade at følge anbefalingen. Et vurderingsskøn er i sagens natur resursekrævende, fordi det forudsætter en stillingtagen til sagens konkrete omstændigheder. Det kan være fristende at etablere faste rammer for skønnet udøvelse som fx gennem beslutningsunderstøttende algoritmer. Som et eksempel på dette kan vi forestille os en regel med følgende variabler: 1) arbejdsløshedsperiode, 2) uddannelsesniveau, 3) alder. Variablerne bruges i en model, hvor arbejdsløshed i x-periode, uddannelse på x-niveau og x-alder indebærer en bestemt retsfølge. Denne tilgang er praktisk, fordi den sætter skønnet under faste rammer, men den vil også være ulovlig, såfremt den tilsidesætter den konkrete og af lovgiver forudsatte stillingtagen til enkeltsagens forhold. Forvaltningen må nemlig ikke ”sætte skøn under regel”, jf. retsgrundsætningen af samme navn (forbuddet mod at sætte skøn under regel). Man må således heller ikke lave en beslutningsunderstøttende løsning i form af et simpelt excel-ark med en formel, hvor man ved at indtaste forudbestemte kriterier for eller imod et givet resultat vejledes om sagens afgørelse, medmindre sagsbehandleren reelt kun bruger oplysningerne vejledende (Motzfeldt, Ullits og Kjellerup, 2020: 175).

Begrundelse

Processgennemsigtighed er en forudsætning for overhovedet at vide, hvad sagsbehandleren tillagde vægt som væsentligt under afgørelsens udarbejdelse, herunder om skønnet er lovligt forvaltet uden inddragelse af usaglige hensyn. Begrundelsespligten stiller netop krav til sagsbehandleren om gennemsigtighed ved at meddele borgeren alle relevante mellemregninger, der ledte frem til den trufne afgørelses resultat. Det gælder også, hvor en beslutningsunderstøttende algoritme er anvendt til dele af sagsbehandlingen.

Det følger af begrundelsespligten i forvaltningslovens § 22, at afgørelser, der ikke giver borgeren fuldt medhold, skal begrundes. Formålet er at fremme borgerens tillid til myndigheden, når borgeren kan se de mellemregninger, som sagsbehandleren inddrog til støtte for resultatet. Derudover er formålet med begrundelsespligten også at øge grundigheden i behandlingen af enkeltsager, da en åbenlyst forkert, mangelfuld eller måske manglende begrundelse i en afgørelse kan få ansættelsesmæssige følger for den pågældende sagsbehandler.

Begrundelserne skal også sikre kvaliteten på hele sagsområdet, idet begrundelserne kan bruges til at koordinere og tilrettelægge praksis. Det er nemlig ikke kun overfor borgeren, men også overfor sagsbehandlerens overordnede, at begrundelsen skal forklare, hvorfor afgørelsen fik netop dette indhold. En grundig begrundelse kan derfor danne præcedens for fremtidige sager. Reglerne om begrundelse varetager således en række hensyn, alle under den overordnede paraply, der vedrører hensynet til borgernes retssikkerhed, ved at give borgerne mulighed for at forholde sig til det grundlag, som afgørelsen er truffet på, og dermed også mulighed for, at en eventuel klage kan ske på et kvalificeret grundlag (Næsborg-Andersen, 2015: 98-106).

For at kunne opfylde formålene med en begrundelse skal den indeholde de retsregler samt relevante oplysninger og vurderinger, som sagsbehandleren har tillagt vægt, og hvis sagsbehandleren har udøvet et vurderingsskøn, skal også dette begrundes. Hvor udførlig begrundelsen skal være, afhænger af det pågældende retsområde, afgørelsens kompleksitet og konsekvenserne for borgeren, idet der stilles større krav til begrundelsens udførlighed ved mere komplekse og/eller indgribende afgørelser, herunder hvis faktum er omtvistet eller bestridt. Det vil dog aldrig være tilstrækkeligt at anføre ”ud fra et samlet skøn”. Derudover har man som minimum pligt til at forholde sig udtrykkeligt til partens bemærkninger, påstande og eventuelle indsigelser, herunder en indsigelse om, at den systemgenererede vurdering er forfæjlet (se også Folketingets Ombudsmand, 2021).

Algoritmer

I denne artikel anvendes begrebet algoritme synonymt med kunstig intelligens (AI) og *machine learning* og således som et paraplybegreb. Vi finder det ikke nødvendigt at redegøre nærmere for de mange forskellige tilgange, der findes til prædikative analyser som fx *deep learning*, neurale netværk, *reinforcement learning*, *supervised learning*, *unsupervised learning* og *natural language processing* eller at dvæle dybere ved de forskellige teknikker. Det er dog væsentligt at huske på – som også EU-kommissionens *Forslag til AI-forordning* (COM(2021) 206 final) gør – at AI grundlæggende er ”software, der ... kan, ud fra et givent sæt af menneskeligt-definerede formål, generere output”.

Algoritmer inkluderer således forskellige teknikker og værktøjer, der gennem matematiske algoritmer baseret på akkumulerede data sætter algoritmen i stand til at generere vurderinger på opstillede problemer. Algoritmer bruges i forskellige myndighedssammenhænge til fx at oversætte sprog i chatrobot på en myndighedsside, generere undertekster i videoer ud fra tilgængelighedshensyn eller blokere e-mailspam.

Algoritmer kan betragtes som sandsynlighedsstatistik, hvor sandsynligheden for et givent resultat genereres ud fra et stort datamateriale som fx afgørelses-tekster, mails mv. Det fungerer sædvanligvis på den måde, at algoritmen kodes til at søge i udvalgte data, såkaldt ”træningsdata”, med henblik på at generere en model, som, når den anvendes i ”virkelige” data (dvs. ikketræningsdata), kan genkende mønstre og korrelationer, som modellen ikke tidligere har set, men som den gennem træningen nu kan kvalificere i en bestemt retning.

Algoritmer ses i dag appliceret i det juridiske arbejde, fx i Miljøstyrelsen, hvor et AI-program skal ”identificere virksomheder, hvor der kan være brug for en særlig myndighedsindsats i form af vejledning og tilsyn” (Andersen, 2019). Andre myndigheder, fx Nævnenes Hus og Ankestyrelsen, er i gang med at inddrage algoritmer til at forbedre kvaliteten i afgørelserne og reducere sagsbehandlingstiden. I alle disse projekter påtænkes algoritmen anvendt som et beslutningsunderstøttende værktøj.

Kendte risici ved brug af algoritmer

De mulige gevinster ved beslutningsunderstøttende algoritmer er italesat som blandt andet en mere effektiv, målrettet og fleksibel borgerservice (se fx Microsoft Danmark og Lead Agency, 2021: 18).

Disse gevinster må dog afvejes overfor andre diskurser, herunder risikoen for såkaldt *automation bias*. Begrebet dækker over de situationer, hvor mennesket stoler blindt eller i hvert fald mere på computerens resultat, end resultatet egentlig kan holde til. Det er altså bias, der gør, at man fejltolker maskinens træfsikkerhed. Litteraturen skelner imellem forskellige typer af mangelfuld brug af beslutningsstøttende algoritmer: *automation bias*, dvs. ukritisk accept af algoritmens resultat; *selective adherence*, dvs. man er mere tilbøjelig til at godtage resultater, som understøtter de forestillinger, beslutningstageren har i forvejen, end resultater som modsiger disse (Alon-Barkat og Busuioc, 2022); og *complacency*, dvs. beslutningstager overser åbenlyse advarsler eller fejlkilder, fordi man stoler på, at computeren har styr på situationen (Wickens et al., 2015). Introduktionen af beslutningsunderstøttende algoritmer kan altså medføre nye typer af fejlkilder i interaktionen imellem beslutningstager og den algoritme, som egentlig har til formål at gøre en beslutning mere korrekt. Disse fejlkilder vil her under et blive omtalt som *automation bias*, idet dette synes at være det gennemgående begreb i forskningen. Fx definerer Manzey, Reichenbach og Onnasch (2012) *automation bias* som en eller flere af tre mulige scenarier: a) manglende opmærksomhed på at efterprøve faktiske oplysninger, b) aktiv fravælgelse af modsigende informationer eller resultater fra algoritmen og c) uopmærksomhed på modsigende informationer eller resultater.

Automation bias er især blevet påvist i situationer, hvor der skal træffes vigtige beslutninger under tidspres, såsom hos piloter eller medicinsk personale, men kan også observeres i andre situationer (Adya og Philips-Wren, 2020), fx blandt administrativt personale (Alon-Barkat og Busuioc, 2022), både når der skal træffes store komplekse beslutninger og mere simple enkeltbeslutninger (Lyell og Coiera, 2017).

Forskningen viser, at beslutningstagerens individuelle erfaringer har betydning, idet erfarne piloter havde mere held med at spotte en åbenlys fejl i navigationssystemet, end deres mindre erfarne kollegaer (de Boer, Heems og Hurts, 2014). Derudover spiller mange andre faktorer ind på alvoren og omfanget af en eventuel automation bias. Stressede beslutningstagere er fx mere tilbøjelige til at acceptere stereotyper og stereotypiske ræsonnementer (Adya og Wren, 2020), ligesom designet af brugerfladen på en beslutningsunderstøttende algoritme kan have indflydelse (Davis, 1989).

Endelig spiller den information, som beslutningstageren modtager fra den beslutningsunderstøttende algoritme, en rolle. Forsøg har således vist, at både manglende information og uoverskuelig eller overflødig information øger automation bias, mens kort og præcis information mindsker den (Wright et al., 2016). Hvis beslutningstageren hurtigt og enkelt kan gennemskue, hvilke faktorer algoritmen har lagt vægt på, vil beslutningstageren således være mindre tilbøjelig til at acceptere en åbenlys forkert anbefaling og mere tilbøjelig til at vurdere, hvorvidt algoritmen er nået frem til det rigtige resultat.

En anden kendt risiko ved brugen af algoritmer er det allerede nævnte black box-paradoks. Paradokset består i, at algoritmen fungerer som en sort (uigenkendselig) boks, der får en mængde input i form af variabler og derefter kommer med et output uden at fortælle, hvordan den kom frem til dette (Pasquale, 2015).

Når den idealtypiske sagsbehandler træffer afgørelser ved brug af juridisk metode uden inddragelse af en algoritmes beregninger, kan afgørelsens anførte resultat begrundes. Det vil således kunne efterprøves, om den anførte begrundelse er korrekt, ved at sammenholde begrundelsen med sagens noter og øvrige journaliserede dokumenter. Algoritmer er derimod outputfokuserede, hvilket betyder, at de er fokuserede på at angive en løsning og ikke på begrundelsen for den valgte løsning. De kan således ikke redegøre nærmere for de intervariable sammenhænge og heller ikke forklare, hvorfor forholdet mellem en inputvariabel og en outputvariabel er fundet kausalt. Altså hvorfor inputvariablen $y + x$ resulterer i outputtet a . I overført betydning svarer det til, at sagsbehandleren inddrager variablerne 1) uddannelsesmæssig baggrund, 2) bopælsadresse og 3) erhvervs erfaring som begrundelse for et bestemt output (retsfølge), men er ude

af stand til at redegøre nærmere for kausaliteten mellem disse intervariable sammenhænge og outputtet.

Beslutningsunderstøttende algoritmers indflydelse på sagsbehandlerens skøn og begrundelse

Det er væsentligt indledende at bemærke, at ASTAs risikoscore har karakter af en procesledende beslutning, og at der for sådanne beslutninger ikke gælder et begrundelseskrav. Tilsvarende skal myndigheder ikke begrunde en beslutning om at gennemføre partshøring. ASTAs risikoscore er, ligesom en partshøring, et processuelt skridt taget med henblik på at kvalificere sagens videre forløb, men derimod ikke – hvilket er afgørende – en oplysning, der i sig selv fastlægger den lediges endelige retstilling. Ikke desto mindre er det relevant at undersøge, om det ville være muligt for sagsbehandlerne at begrunde ASTAs risikoscore i en situation, hvor denne score påtænkes anvendt i en afgørelse overfor den ledige. Med dette forbehold, og med forbehold for at udviklerne og myndigheden ville have tilrettelagt processerne anderledes, såfremt der havde været tale om en decideret afgørelsesunderstøttende løsning, og at ASTA var et pilotprojekt og ikke endeligt implementeret (og nu droppet grundet manglende lovhjemmel (Datatilsynet, 2022)), gennemføres i det følgende en analyse af den mulige påvirkning af sagsbehandlerens skøn, som anvendelsen af ASTA kan afstedkomme, samt af de problemstillinger, et sådant system kan medføre i forhold til begrundelseskravet.

Det er op til den enkelte sagsbehandler, om de vil godtage risikoscoren, eller om de vil ændre den ved at lave et notat i sagen. Der kan således være tale om, at sagsbehandleren accepterer scoren både uden at tænke videre over det eller ved at undersøge, hvordan algoritmen har fundet frem til scoren (hvis muligt), og derefter godtager den. Sagsbehandleren kan på samme måde afvise resultatet blankt uden videre undersøgelse eller efter en nærmere vurdering af, hvorvidt de forskellige variabler er blevet vægtet korrekt. I de tilfælde hvor sagsbehandleren vurderer scorens validitet efter en nærmere undersøgelse af vægtningen af de forskellige variabler, vil sandsynligheden for, at der er tale om automation bias, alt andet lige være mindre, end i de tilfælde hvor sagsbehandleren accepterer eller afviser resultatet uden nærmere overvejelser.

Automation bias

Det er et gennemgående træk i Flügge et al.s (2022) undersøgelse af sagsbehandlerens brug af ASTA, at sagsbehandlerne sjældent undersøger vægtningen af de forskellige variabler, heller ikke i de tilfælde hvor de er uenige i scoren. Dette skyldes flere forhold, heriblandt at sagsbehandlerne har svært ved at se

relevansen af nogle af variablerne, og at andre opfattes som direkte misvisende. Det omfatter blandt andet variabelen om, hvorvidt jobsøgeren har opdateret sit elektroniske CV, som bliver inddraget på et tidspunkt, hvor jobsøgeren endnu ikke er blevet instrueret i at gøre dette.

Sagsbehandlerne påpeger også, at de ikke er blevet instrueret eller vejledt i, hvordan de skal vurdere vægtningen af de forskellige variabler. Derfor er motivationen for at bruge arbejdstid på dette ikke stor (Flügge et al., 2022: 17-18). Fordelingen af jobsøgere i de tre farver opleves i det hele taget som værende uigennemsigtig. Mange jobsøgere placeres især i det gule felt, selvom sagsbehandlerne vurderer dem som værende grønne. Samtidig har sagsbehandlerne ikke nogen viden om, hvad ASTAs fejlmargen er, og er derfor utrygge ved at stole for meget på risikoscorerne (Flügge et al., 2022: 13-19).

Fra ASTA-rapporten udtaler en sagsbehandler skepsis om det beslutningsunderstøttende systems præcision:

S3 nævner eksempelvis, at hvis jobcentret modtog 300 nyledige, så oplever S3, at ASTA kategoriserer ca. 20-25 som værende grønne eller røde. Og altså 250 ud af 300 nyledige som gule. Det stemmer i sagsbehandlerens optik ikke overens med sagsbehandlerens opfattelse af virkeligheden. Kort sagt er der for få grønne og røde og for mange gule. Som S3 siger det ”Der er en kæmpe overvægt af de gule, og hvis de gule er hverken eller, så kan jeg ikke bruge systemet til noget. Der skal være flere, der kommer ud i yderkanterne. Grøn eller rød. Derfor synes jeg ikke systemet kan bruges til særlig meget” (Flügge et al., 2022: 16).

Udtalelsen illustrerer, at sagsbehandleren inddrager sin faglighed og erfaring som målestok for løsningens præcision og fremhæver en vis bias – ikke i løsningen, men hos sagsbehandleren. Sagsbehandleren er vant til en bestemt outputmassesammensætning og har indrettet sig herefter, således at den nu opleves som det rigtige. En menneskelig bias, der omtales *present bias* (Coglianese og Lay, 2022: 1295). Da ASTA afviger fra den hidtidige sagsmassesammensætning og er ude af stand til at redegøre for kausaliteten bag de intervariable sammenhænge og dets output, står sagsbehandlerne overfor et valg mellem enten at følge egne erfaringer eller at udmønte ASTAs maskinelle vurderinger, der indimellem adskiller sig væsentligt fra hidtidige erfaringer hos sagsbehandlerne.

En omvendt problemstilling forekommer, hvis sagsbehandlerne stoler for meget på den beslutningsunderstøttende løsning og helt tilsidesætter deres vurderingsskøn. I forhold til Manzey, Reichenbach og Onnaschs (2012) definition ville der her være tale om automation bias både i form af manglende efterprøvelse af fakta og ved en aktiv fravælgelse af modstridende resultater. Hvis sags-

behandlerne reelt aldrig undersøger de udslagsgivende variabler, kan de ikke vide, om de har overset modstridende eller direkte forkerte resultater.

Som sagsbehandlingen er tilrettelagt, forekommer det dog ikke i ASTA-sagen muligt at undersøge, hvilke variabler der har været udslagsgivende for ASTAs scoring (hvorvidt det er teknisk muligt diskuteres nedenfor). Det kan tilskrives, som beskrevet af Wright et al. (2016), at mængden af informationer til sagsbehandleren er for stor og indeholder for mange overflødige oplysninger, hvilket øger risikoen for automation bias. Den ene sagsbehandler påpeger, at man som sagsbehandler "[s]let ikke har tid til at gennemgå 50 variabler" (Flügge et al., 2022: 14). Når risikoen for automation bias øges, stiger sandsynligheden givetvis også for, at sagsbehandlerne tilsidesætter deres fagprofessionelle vurdering og i uhensigtsmæssig grad følger værktøjets anbefaling (således Eubanks, 2018). Der ses dog ikke eksempler på dette i ASTA-rapporten, måske fordi den blev lavet på baggrund af et begrænset antal data, da ASTA stadig var i testfasen og dermed ikke var integreret i sagsbehandlerens hverdag.

Når sagsbehandlerne således ikke forstår algoritmens output og samtidig ikke har indseende i mellemregningerne, så er der, som skitseret ovenfor, groft sagt to mulige sagsudfald. At algoritmen aktivt anvendes til trods for sine blinde vinkler, eller, som tilfældet forekommer at være i ASTA-sagen, at algoritmeversionen hos sagsbehandlerne bliver en overskyggende faktor og bevirker, at algoritmen de facto ikke anvendes (de Boer, Heems og Hurts, 2014; Dietvorst, Simmons og Massey, 2015). Det er i førstnævnte situation, at algoritmen vil kunne påvirke sagsbehandlerens skøn, og konsekvenserne ved ukritisk at lægge algoritmens vurdering til grund vil være, at skævheder i data (bias) afspejles i afgørelsen. Når modellen eksempelvis trænes på tidligere praksis og statistik og heri finder en række sammenhænge, der senere lægges til grund af sagsbehandleren, reproduceres disse sammenhænge, herunder også en iboende bias i de korrelationer, som algoritmen har identificeret og tillagt vægt. Som illustreret ved den nu forladte såkaldte Gladsaxe-model, der var en beslutningsunderstøttende algoritme til tidlig opsporing af udsatte børn i Gladsaxe kommune, trænede denne på statistik og andre historiske data (Næsborg-Andersen og Motzfeldt, 2019). De anvendte statistiske data til træning af Gladsaxe-modellen viste, at der er flere udsatte børn fra fattige familier og familier med en ikkevestlig baggrund, hvilket øgede sandsynligheden for, at algoritmen udpegede et større antal af sådanne familier i kategorien "risiko for udsatte børn" og færre velstående familier af dansk eller vestlig oprindelse end de konkrete omstændigheder til sagde. Dataskævheder i en beslutningsunderstøttende algoritme er selvstændigt problematiske, men vil dog kunne opfanges og adresseres af sagsbehandleren, mens kombinationen af dataskævheder og automation bias er retssikkerheds-

mæssigt kritisk, fordi sagsbehandleren grundet magtesløshed reelt sættes udenfor indflydelse. Når sagsbehandleren ukritisk følger den algoritmiske vurdering, er algoritmen de facto ikke længere understøttende, men outputdeterministisk.

Black box-paradokset

Fra bias til en anden udfordring: Black box-paradokset beskriver som nævnt de situationer, hvor man oplever, at algoritmen er uigennemsigtig. Dette ses der klare eksempler på i rapporten om ASTA. En sagsbehandler udtaler direkte: ”Jeg ved ikke, hvor præcis ASTA er. Den giver ikke nogen forklaring af, hvor god den er til at ramme rigtigt, og hvem siger, hvad der er rigtigt? Jeg kender ikke variablerne, og der er ikke forklaring af dem” (Flügge et al., 2022: 18).

At algoritmer som altovervejende hovedregel som nævnt er outputfokuserede betyder, at de er fokuserede på at angive en løsning. De er derimod ikke designet til at forklare, hvorfor lige netop en specifik løsning er fundet optimal. ASTA er ingen undtagelse. De testende sagsbehandlere er for det meste oplyst om, hvilke variabler ASTAs scoring *kan* baseres på, men får tilsyneladende ingen oplysninger om, hvordan variablerne *er* vægtet i de konkrete sager. En anden sagsbehandler beskriver således sin oplevelse af ASTAs risikoscore: ”Risikovurderingen er en lang liste af spørgsmål og svar. Det er alt muligt om deres liv. Deres liv på et stykke papir. Og det er helt umuligt at danne sig et overblik over eller forstå dem alle” (Flügge et al., 2022:18).

Der er altså en tydelig oplevelse blandt sagsbehandlerne af, at ASTAs algoritme er uigennemsigtig, og der er ingen forståelse af, hvordan de forskellige variabler bliver vægtet. Denne oplevelse skal sammenholdes med de juridiske krav til begrundelse og skøn. Hvis ASTAs score var en medvirkende faktor i en afgørelse om at placere en jobsøger i en bestemt risikokategori med tilhørende krav og tilbud, ville der være tale om en væsentlig oplysning. Dermed stilles også krav om at redegøre nærmere for dette output, fx ved at beskrive kausaliteten mellem algoritmens intervariable sammenhænge og afgørelsens resultatet (fx [angivne alder] + [helbredssituation] + [indkomst] + [uddannelsesmæssig baggrund] = [output]). De bærende hensyn bag beslutningen skal nemlig angives, og det er ikke tilstrækkeligt at anføre, at ASTAs score er indgået i det samlede skøn.

ASTA anvender som bekendt 50 variabler, mens et menneskes korttidshukommelse antages at være begrænset til at operere med omkring fire variabler (Cowan, 2001). Sagsbehandleren vil således ikke være i stand til at dechiffrere algoritmens samlede beregning (output), når denne, som i ASTA-sagen, er dannet ud fra et større antal vægtede variabler, og vil heller ikke være i stand til

overfor borgeren at forklare baggrunden for, at man har valgt at følge anbefalingen og træffe afgørelse i henhold til denne.

Når sagsbehandlerne således ikke kognitivt eller via systemteknisk understøttelse tilvejebringes en form for algoritmisk transparens, er det ikke muligt at begrunde ASTAs vurdering og derved ikke muligt for borgeren at efterprøve sin retsstilling, fx om algoritmen har undladt at lægge vægt på en væsentlig oplysning. Det samme gælder for en eventuel klageinstans, der skal vurdere en algoritmisk understøttet afgørelses substantielle lovlighed. Uden indsigt i vægtningen af variabler og inddragne hensyn er det ikke muligt. *Paradokset* i black box er således, at algoritmen på den ene side kan behandle langt større datamængder mere effektivt end sagsbehandleren, men at der på den anden side ikke kan støttes ret på algoritmens ubegrundede outputs, idet afgørelser legitimeres via begrundelsen.

Konklusion

Det er evident og ukontroversielt, at beslutningsunderstøttende algoritmer påvirker sagsbehandlerens kvalificering af sagen. Formålet med beslutningsunderstøttende algoritmer er netop at effektivisere og forbedre sagsbehandlingen. Men det er essentielt at overveje de retssikkerhedsmæssige konsekvenser ved en sådan algoritmisk påvirkningsfaktor.

Via ASTA-rapportens data har vi undersøgt to problemstillinger i forhold til beslutningsunderstøttende algoritmer, nemlig automation bias og black box-paradokset, som på hver sin måde vil kunne påvirke sagsbehandlerens skøn og begrundelse. Hvad angår automation bias, kan vi overordnet konkludere, at algoritmeaversionen mod ASTA har været så markant, at nogle sagsbehandlere slet ikke anvender output fra ASTA. Herved opstår der ikke algoritmisk bias, da output ikke har mulighed for at trænge igennem til den retlige kvalificering. Vi kan om bias mere generelt konstatere, at dataskævheder i en beslutningsunderstøttende algoritme er problematiske, men dog vil kunne opfanges af sagsbehandleren som en form for ”biasfilter” (givet at det er kognitivt muligt, og at sagsbehandleren er klædt på til dette). Kombinationen af dataskævheder og automation bias er derimod langt mere kritisk, fordi sagsbehandleren sætter sig selv udenfor indflydelse, hvorved den algoritmiske understøttelse de facto bliver outputdeterministisk. Under sådanne omstændigheder opstår der således en risiko for, at algoritmen får en utilsigtet indflydelse på sagsbehandlerens skøn. I forhold til black box-paradokset var dette fremtrædende i ASTA-rapporten, hvilket skyldtes algoritmens kompleksitet (antallet af variabler og intervariable sammenhænge). Sagsbehandlerne kunne således hverken kognitivt eller via systemteknisk understøttelse i øvrigt få indsigt i baggrunden for ASTAs vur-

dering, og det retlige krav om begrundelse ville således ikke kunne opfyldes, havde man inddraget vurderingen i en afgørelse.

Der er ikke et juridisk krav om, at enhver oplysning skal begrundes, og heller ikke krav om at informere om samtlige variabler, algoritmetype mv. Det afgørende må være, at sagsbehandleren kan begrunde outputtet, hvilket muligvis forudsætter information om visse variabler mv. Samlet set har black box-paradokset dog påvirket sagsbehandlerne i ASTA-sagen, ikke i forhold til begrundelsespligten, men ved at medvirke til den udbredte algoritmeaversion blandt dem. For en beslutningsunderstøttende algoritme skal understøtte, hvilket forudsætter et samarbejde mellem sagsbehandler og algoritme, så sagsbehandlerne får indsigt i vægtningen af variabler og tillagte hensyn. Selvom algoritmer kan være effektive og tilvejebringe bedre sagsbehandling, skal sagsbehandlerne være kompetente til at inddrage algoritmens output i deres skøn og legitimere afgørelsen i begrundelsen. Kun således overholdes retsstatens krav til en objektiv og saglig sagsbehandling.

Litteratur

- Adya, Monica og Gloria Phillips-Wren (2020). Stressed decision makers and use of decision aids: A literature review and conceptual model. *Information Technology and People* 33 (2): 710-754.
- Alon-Barkat, Saar og Madalina Busuioc (2022). Human-AI interactions in public sector decision making: “Automation bias” and “selective adherence” to algorithmic advice. *Journal of Public Administration Research and Theory* 33 (1): 153-169.
- Andersen, Søren Stig (2019). eForvaltning på miljøområdet. *Tidsskrift for miljø* 8: 259-269.
- Bryman, Alan (2016). *Social research methods*. Oxford: Oxford University Press.
- Bønsing, Sten (2018). *Almindelig forvaltningsret*, 4 udgave. København: Jurist- og økonomiforbundets forlag
- Coglianesi, Cary og Alicia Lay (2022). Algorithm vs. algorithm. *Duke Law Journal* 72: 1281.
- Cowan, Nelson (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral Brain Science* 24 (1): 87-114.
- Datatilsynet (2022). Udtalelse fra Datatilsynet: Kommuners hjemmel til AI-profileeringsværktøjet Asta. <https://www.datatilsynet.dk/afgoerelser/afgoerelser/2022/maj/udtalelse-vedroerende-kommuners-hjemmel> (4. januar 2023).
- Davis, Fred D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 13 (3): 319-340.

- de Boer, Robert J., Wijnand Heems og Karel Hurts (2014). The duration of automation bias in a realistic setting. *The International Journal of Aviation Psychology* 24 (4): 287-299.
- Dietvorst, Berkeley J., Joseph P. Simmons og Cade Massey (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144 (1): 114-126.
- Eubanks, Virginia (2018). *Automating inequality*. New York: St. Martin's Press.
- Fenger, Niels (red.) (2018). *Forvaltningsret*. København: Jurist- og Økonomforbundets Forlag.
- Flügge, Asbjørn Ammitzbøll, Naja Holten Møller, Thomas T. Hildebrandt og Henrik Palmer Olsen (2022). *Er du grøn? Algoritmer til beslutningsstøtte i det offentlige. En kvalitativ undersøgelse af sagsbehandleres praksis og brug af ASTA til profilering af nyledige dagpengemodtagere*. Københavns Universitet, Det Naturvidenskabelige Fakultet.
- Folketingets Ombudsmand (2021). *Vurderingsstyrelsens begrundelse og stillingtagen til partsanbringender i sager om omvurdering efter vurderingsloven*. FOB 2021-25. https://www.ombudsmanden.dk/find/udtalelser/beretningssager/alle_bsager/2021-25/ (28. april 2023).
- Forslag til AI Forordning (COM(2021) 206 final (2021). <https://eur-lex.europa.eu/legal-content/DA/TXT/?uri=COM:2021:206:FIN> (4. januar 2023)
- Loiborg, Emilie (2020). Om den forvaltningsretlige ramme for overladelse af forvaltningsopgaver til systemer baseret på kunstig intelligens. *Ugeskrift for Retsvæsen* (U.2020B.64).
- Loiborg, Emilie og Jakob Klaris Søgaard (2018). Offentlige myndigheders delegation af myndighedsudøvelse ved digitalisering af den offentlige forvaltning. *Ugeskrift for Retsvæsen* (U2018B.90).
- Lyell, David og Enrico Coiera (2017). Automation bias and verification complexity: A systematic review. *Journal of the American Medical Informatics Association* 24 (2): 423-431.
- Manzey, Dietrich, Juliane Reichenbach og Linda Onnasch (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making* 6 (1): 57-87.
- Meilvang, Marie Leth og Anne Marie Dahler (2022). Decision support and algorithmic support: The construction of algorithms and professional discretion in social work. *European Journal of Social Work*. <https://doi.org/10.1080/13691457.2022.2063806>
- Microsoft Danmark og LEAD agency (2021). *Kunstig intelligens i den offentlige sektor*. https://mst.dk/media/173776/microsoft_rapport_raekkefoelge.pdf (28. april 2023).

- Motzfeldt, Hanne Marie (2015). Den danske ombudsmands krav om værdibaseret design af den digitale forvaltning. *Nordisk Administrativt Tidsskrift* 92 (3): 8-23.
- Motzfeldt, Hanne Marie (2017). The Danish principle of administrative law by design. *European Public Law* 23 (4): 739-754.
- Motzfeldt, Hanne Marie (2020). Machine learning og forvaltningens skønsudøvelse. *Juristen* 4: 140-147.
- Motzfeldt, Hanne Marie og Azad Taheri Abkenar (2019). *Digital forvaltning: Udvikling af sagsbehandlede løsninger*. København: DJØF Forlag.
- Motzfeldt, Hanne Marie og Ayo Næsborg-Andersen (2018). Developing administrative law into handling the challenges of digital government in Denmark. *Electronic Journal of E-Government* 16 (2): 136-146.
- Motzfeldt, Hanne Marie, Jøren Ullits og Jens Kjellerup (2020). *Fra forvaltningsjurist til udviklingsjurist: Introduktion til offentlig digitalisering* (1. udg.). København: DJØF Forlag.
- Motzfeldt, Hanne Marie, Jøren Ullits, Emilie Loiborg og Jens Kjellerup (under udg.). *Fra forvaltningsjurist til udviklingsjurist: Introduktion til offentlig digitalisering* (2. udg.). København: DJØF Forlag.
- Mørup, Søren Højgaard, Jens Garde, Jørgen Albæk Jensen, Orla Friis Jensen, Helle Bødker Madsen, Karsten Revsbech og Ole Terkelsen (2022). *Forvaltningsret – almindelige emner* (7. udg.). København: Jurist- og Økonomforbundets Forlag
- Nielsen, Stine Piilgaard Pøner og Ole Hammerslev (2022). Velfærdsstatens regulering af ungdomshjemløshed: Retlige handlemuligheder og deres potentielle effekt. *Retferd. Nordic Journal of Law and Justice* 172 (2): 11-22.
- Næsborg-Andersen, Ayo (2015). *Human rights in national administrative law: Dissemination of knowledge of human rights through administrative decisions*. København: DJØF Forlag.
- Næsborg-Andersen, Ayo og Hanne Marie Motzfeldt (2019). Different aspects of transparency in digital government: The Danish case. *European Conference on Digital Government*, 81-87, VIII-IX.
- Olsen, Henrik Palmer, Jacob Livingstone Slosser, Thomas Troels Hildebrandt og Cornelius Wiesener (2019). *What's in the box? The legal requirement of explainability in computationally aided decision-making in public administration*. Social Science Research Network.
- Pasquale, Frank (2015). *The black box society: The secret algorithms that control money and information*. Cambridge: Harvard University Press.
- Pound, Roscoe (1910). Law in books and law in action. *American Law Review* 44 (1): 12-36.
- Rønsholdt, Steen (2018). *Forvaltningsret – retssikkerhed proces sagsbehandling*. København: Karnov Group.

- Stausholm, Nick Storm (2015). *Skøn under regel*. København: Karnov Group.
- Wickens, Christopher D., Benjamin A. Clegg, Alex Z. Vieane og Angelia L. Sebok (2015). Complacency and automation bias in the use of imperfect automation. *Human Factors* 57 (5): 728-739.
- Wisborg, Thea Johanne Raugland (2022). *Fuldautomatiserede afgørelser i dansk forvaltning – en retlig analyse med udgangspunkt i databeskyttelsesforordningens artikel 22*. Aarhus: Aarhus Universitet
- Wright, Julia. L., Jessie Y. C. Chen, Michael J. Barnes og P.A. Hancock (2016). The effect of agent reasoning transparency on automation bias: An analysis of response performance, pp. 465-477 i Stephanie Lackey og Randall Shumaker (red.), *Virtual, augmented and mixed reality*. Cham: Springer.