

Kunstig Intelligens og Medicinsk Etik: Tilfældet *Watson for Oncology*

Ezio Di Nucci, Aaro Tupasela & Rasmus Thybo Jensen

Indledning

Der kan ikke herske tvivl om at medicinstuderende og generelt studerende på sundhedsuddannelserne er motiverede af at udsigten til at kunne yde hjælp til folk der har behov for det. Men skal vi være ærlige, må vi anerkende at udsigten til et velbetalt, sikkert job også er en motivator for f.eks. medicinstuderende. Det er derfor at kunstig intelligens (*Artificial Intelligence* eller *AI*) inden for sundhedsvæsenet bør skræmme dig, om du er en af disse studerende: AI er ude efter dit job.

I dette kapitel diskuterer vi *IBM Watson for Oncology* (kaldet *Watson* fremover) som et casestudie i den fremadstormende medicinske AI. Vi vil analysere nogle af de mest interessante etiske og filosofiske spørgsmål som brugen af AI i sundhedsvæsenet rejser med et særligt fokus på *Watson*.

Watson er et såkaldt “beslutningsunderstøttende [*decision-support*] system der rangordner kræftbehandlingstilbud” (Di Nucci 2019: 1) ved hjælp af maskinlæringsalgoritmer, dvs. computerprogrammer der, ifølge computervidenskabsfolk, er i stand til “på egen hånd at regne den ud, ved at lave slutninger fra data” (Domingos 2015: xi).

Så du kan godt doble din frygt op: De grådige maskiner er ikke kun ude efter dit job, de er også ude efter deres egne programmørers job. De synes i det hele taget ikke at helme, før de har overtaget

hele verden, hvilket præcist er det som teknofober og deres ekstremistiske venner *teknopokalytikerne* frygter (se f.eks. Tegmark 2017, særligt kapitel 1).¹

Hvordan virker Watson? Baseret på de nyeste medicinske forskningspublikationer og patientens journal frembringer Watsons algoritmer kræftsbehandlingsanbefalinger som kræftlægen kan gennemgå og bruge i sin konsultation med patienten. Algoritmerne er udviklede af IBMs programmører i samarbejde med kræftlæger fra *Memorial Sloan Kettering Cancer Center* ((MSK) i New York.

Her lidt flere detaljer om hvordan Watson fungerer, hentet fra en nylig analyse af Rosalind McDougall i *Journal of Medical Ethics*:

Watson for Oncology is designed and marketed as a tool for clinicians, to assist them to ‘zero in on the most promising care options’ in an age where the available literature on cancer is huge and fast-moving. The system extracts clinical information from the patient’s medical record, such as gender, age, stage and type of cancer, family history, notes from previous visits, test results and comorbidities. The doctor is prompted to verify this extracted information and add additional relevant information. The information is then analysed, based on the computer’s training by oncologists at Memorial Sloan Kettering Cancer Center in New York. The system accesses over 300 medical journals and over 200 textbooks. According to IBM’s marketing video, Watson ‘identifies a prioritized list of treatment options based on Memorial Sloan Kettering expertise and training, and provides links to supporting evidence’. Watson presents a ranked list of treatment options and a synthesis of the existing published evidence relevant to that clinical situation. The treatment options are divided into three colour-coded sections: green for recommended treatments, amber for treatments to consider and red for treatments that are not recommended. There may be multiple options in each section. The options are ranked based on outcome statistics presented in terms of ‘disease-free survival’. For each treatment option, there are two literature tabs available to the clinician. One tab gives links to literature that supports that treatment option, identified by Memorial Sloan Kettering clinicians. The other tab gives links to Watson-identified literature relevant to that clinical situation. Information on toxicities associated with the treatment (such as vomiting, anaemia, diarrhoea and so on) is also available to clinicians. The system also has the capacity to be customised to the specific

¹ Det findes også, skal det siges, mere optimistiske stemmer som f.eks. Eric Topols *Deep Medicine*.

geographic context where it is being used. For example, local clinical guidelines and availability of drugs are included for clinicians (McDougall 2019: 157).²

Er du ved at være bange? Den form for medicinsk rådgivning og mediciske anbefalinger som patienter plejede at få fra kræftlægerne synes at være blevet uddelegeret til computeralgoritmer, og det endda til intelligente algoritmer som kan programmere sig selv (om vi ellers kan tro på visse computervidenskabsfolk). Umiddelbart kunne det se ud til at vi alle burde være bange: patienterne fordi deres behandling bestemmes af computere og lægerne fordi deres jobs bliver outsorceret.

Automatiseringen (Wajcman 2017, Susskind & Susskind 2015) vil være nået hele vejen rundt når selv prestigefyldte og velbetalte jobs, som f.eks. lægens eller advokatens, mistes til maskinlæringsalgoritmer.³

Grundlag for frygt? Uddelegering og præstationsevne

Vi vil mene at det ovenstående noget dystre scenarie hviler på forhastede konklusioner. Der er rigtignok alvorlige etiske og filosofiske spørgsmål der bør rejses, men først har vi brug for en ordentlig analyse af hvad Watsons opgaver er. Først derefter kan vi bedømme sådanne teknologiske innovationer ud fra et etisk perspektiv.

² IBM's marketingsmateriale kan findes her: https://www.youtube.com/watch?v=8_bi-S0XNPI (sidst tjekket 28.10-2019). I disse populærvidenskabelige artikler findes flere detaljer om Watson:

<http://fortune.com/2018/07/27/ibm-watson-cancer/>

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

<https://www.engadget.com/2017/06/01/ibm-watson-cancer-treatment-plans/?guccounter=1>

³ Det er muligvis en ringe trøst, men det ikke kun lægers jobs der siges at være truede men også f.eks. advokaters (Fry 2018).

Så hvad er det mere præcist som uddelegeres til Watson? Der er f.eks. stor forskel på at uddelegere egentlig beslutningstagen og at uddelegere rådgivning, eller, som det kaldes i denne sammenhæng, 'beslutningsunderstøttelse' til algoritmer. Og givet den ovenstående beskrivelsen af Watsons ser det ud til at det som uddelegeres, er understøttelse af beslutninger snarere end egentlig beslutningstagen. Ideen er at Watson inddrages for at understøtte klinikerens beslutning og ikke træffer beslutninger på vegne af eller som erstatning for kræftlægen. I hvert fald på papiret forbliver beslutningskompetencen og autoriteten altså hos klinikerens sammen med patienten. Men det er værd at bemærke at Watson faktisk er designet til at komme op med løsninger på de opgaver som den siges at skulle understøtte de menneskelige eksperter i at løse. Måden Watson understøtter eksperternes beslutning om hvilke behandlinger der skal anbefales er, i hvert fald delvist, ved at selv at fremkomme med en prioriteret liste over hvilke behandling der bør anbefales. Det betyder at der i teorien er et kontinuum af mulige brug af Watson, som går fra at lade Watson have meget lille indflydelse på klinikerens beslutning og til, i den anden ende af skalaen, at lade Watson få fuld kontrol over hvilken behandling der anbefales til en given patient. Disse mulige brug betyder at vi i praksis konstant står over for et valg om, hvor langt vi er villige til at gå i forhold til at uddelegere opgaver til Watson.

Den basale teoretiske distinktion mellem beslutningstagning og beslutningsunderstøttelse viser at vi har brug for at strukturere vores analyse på en måde der er en anelse mere sofistikeret end den som kendetegnede den overfladiske teknofobiske skepsis vi indledte med.

Lad os prøve at opdele analysen via følgende fire spørgsmål:

- 1) Hvad kunne være potentielle grunde til uddelegere henholdsvis beslutningskompetene og beslutningsunderstøttelse til Watson?

- 2) Hvor god er Watson og hvordan kan vi måle Watsons præstationsevne både i forhold til en absolut målestok og i forhold til menneskelige eksperter?
- 3) Er der risici ved at bruge Watson der er uafhængige af eventuelle problemer med at måle Watsons ydelsesniveau?
- 4) Og endelig: Hvilke opgaver er vi – etisk set – berettigede til at uddelegere til Watson, og hvilke bør vi, igen ud fra et etisk perspektiv, reservere til mennesker?⁴

Spørgsmål #4 er det overordnede spørgsmål som vi skal svare på, når vi forsøger at give en etisk analyse af Watson. Men for at gøre det må vi først adressere de tre øvrige spørgsmål samt en del andre undervejs. Så lad os bare starte med Spørgsmål #1.

Spørgsmål #1: Uddelegering af hvad til hvem og hvorfor?

Som Spørgsmål #1 antyder, så er kontrol en vigtig del af den etiske ligning når det gælder Watson. Beslutninger i sundhedsvæsenet, særligt når de potentielt er spørgsmål om liv eller død, er for vigtige til at overlade kontrollen til algoritmer som Watson. Hvor meget kontrol vi bør bibeholde afhænger af hvor meget vi i en given kontekst kerer os om og bør kere os om det pågældende område. Og vi kerer os om og bør kere os om kræft fordi den slår mennesker ihjel og fordi fremskridt inden for kræftforskning kan redde mange menneskeliv. En nylig statistik fra Storbritannien fortæller at hver anden person i live i dag (i UK) vil få diagnosticeret kræft i løbet af sit liv og at halvdelen af disse vil overleve sygdommen.⁵ Kræftbehandling betyder noget. Derfor

⁴ Læg her mærke til den klassiske etiske distinktion mellem på den ene side det at en handling er moralsk tilladelig, og på den ene side det at en handling er moralsk påbudt eller obligatorisk (se Di Nucci 2018, Kaptiel 3, for en diskussion af denne distinktion).

⁵ De mere præcise tal kan ses her:

<https://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/lifetime-risk>

ønsker vi ikke at miste kontrollen med behandlingen (den menneskelige kontrol vel og mærke).⁶ Overført til vores case, betyder dette ræsonnement at vi bør være tilbageholdende med at lade maskiner overtage kontrollen med hvilke anbefalinger om behandling en patient bør modtage.

En mulig indvending mod denne konklusion er at vi ikke a priori, dvs. på forhånd og uden hensyntagen til faktisk forhold, bør sætte nogen som helst absolut grænse for hvor langt vi kan tillade os at gå i vores uddelegering. Det kunne jo være at tabet af kontrol opvejes af det faktum at AI-systemet har væsentlig højere præstationsevne i forhold til at udføre den relevante opgave sammenlignet med de bedste menneskelige eksperter. Dette er en fair indvending, men som den er formuleret her har den dog, som vi skal i det følgende, brug for at blive nuanceret.

Vi har brug for at se nærmere på hvad det overhovedet vil sige at f.eks. Watson har en bedre præstationsevne end de menneskelige eksperter, eftersom der er mindst to forskellige måder hvorpå sådanne teknologiske systemer kan forbedre præstationsevnen.

Den første måde et system kan forbedre vores præstationsevne er når det system vi uddelegerer en opgave er i stand til at udføre opgaven bedre end mennesket og måske endda bedre end noget menneske nogensinde vil kunne. At uddelegere fordi vi kan opnå en sådan forbedring i præstationsevne vil vi kalde *strategisk uddelegering*. Her uddelegerer vi til nogen eller noget fordi disse antages at udføre opgaven på et højere niveau; de er så at sige de virkelige eksperter. Strategisk uddelegering finder ikke bare sted mellem mennesker og teknologier men i høj grad også mellem forskellige mennesker, som når vi f.eks. lader en blikkenslager reparere rørene i køkkenet.

Med hensyn til Watson så ser vi hvordan IBM fremhæver en specifik strategisk begrundelse for at bruge Watson i deres marketingsmateriale når de skriver følgende: “With the ability to read over

⁶ Se Santoni & Van der Hoven (2018) for en filosofisk analyse af begrebet ”meaningful human control”. Se Grote & Di Nucci (forthcoming) for en mere udførlig diskussion af vigtigheden af begrebet om kontrol i debatter om algoritmisk beslutningstagning.

200 million pages in three seconds, IBM Watson® for Oncology is learning the language of oncology and detecting patterns that humans may not have seen.”⁷ Her synes den bagvedliggende tanke at være at Watson har en evne til at opfange mønstre i data som potentielt overgår selv den dygtigste menneskelige ekspert. Som sådan udgør citatet et eksempel på hvordan den teoretiske ide om strategisk uddelegering kan finde anvendelse i forhold til Watson.

Der findes et alternativet til strategisk uddelegering, nemlig det vi vil kalde *økonomisk uddelegering*. Her finder vi den anden måde vi kan tale om at et system kan forbedre vores præstationsevne. Den optræder når vi uddelegerer en opgave til nogen eller noget ikke fordi vi tror at de kan udføre opgaven bedre end den hidtil er blevet udført, men fordi vi tror at personen eller systemet der uddelegeres til kan udføre opgaven mere ressource-effektivt. Rengøringshjælp kan være et godt eksempel på dette. De som hyrer rengøringshjælp kan være overbeviste om at de selv kunne gøre det ligeså godt (også selvom de faktisk ikke vil kunne det), men hyre hjælp alligevel fordi de finder det mere ressource-effektivt at uddelegere opgaven. Det kunne f.eks. være hvis de derved får mulighed for at arbejde længere tid til en højere løn end den de betaler deres rengøringshjælp.

Helt generelt, så afhænger en strategisk grund til at uddelegere en given opgave ikke nødvendigvis af at man højner kvaliteten af udførelsen af den pågældende opgave. Den afhænger derimod af at man kan spare et givet systems ressourcer (f.eks. en husholding eller et sundhedsvæsen), på en sådan måde at det tillader en optimering af systemets overordnede præstation.

Hvordan finder begrebet om økonomisk uddelegering anvendelse i forhold til vores case? I stedet for at argumentere for at Watson kan gøre ting bedre end kræftlægerne, kunne vi argumentere for at lægerne, som måske er stressede og pressede på tid, kan få tid til mere meningsfulde opgaver ved at

⁷ <https://www.ibm.com/downloads/cas/AONROW12>

overlade til Watson at skanne den nyeste medicinske forskningslitteratur for relevante artikler. På den måde kunne den dyre og begrænsede tid som de menneskelige eksperter har til rådighed udnyttes mere effektivt, f.eks. kunne det være at lægerne ville få mere tid til hver enkelt patient.

Det synes at være en sådan økonomisk grund til uddelegering som IBM giver når de skriver følgende: “Recent studies continue to demonstrate that Watson for Oncology ‘agrees’ with physicians around the world in the vast majority of cases – so experts can focus on what they do best— deliver care” ([IBM 2018](#)).⁸

Ideen er altså, at selv hvis Watsons præstationsevne rent kvalitativt kun er lige så god og ikke bedre end de menneskelige eksperter og der således ikke er nogle strategiske grunde til at uddelegere, så kunne der stadig være uafhængige grunde der taler for at bruge Watson i kraft af ideen om økonomisk uddelegering. Sådanne økonomiske grunde til at uddelegere kunne endda være til stede selv hvis AI-systemet er marginalt dårligere end de menneskelige eksperter, forudsat at brugen af systemet tillader tilstrækkelig stor besparelse enten pengemæssigt eller i forhold til andre ressourcer.

Så snart vi har fået øje på skellet mellem strategisk og økonomisk uddelegering, så ser vi også at et teknologisk systems isolerede præstationsevne er afgørende men ikke kan udgøre hele historien når vi skal afgøre om der er grund til at uddelegere en opgave. Der vil altid være en kontekst der spiller ind, nemlig den velkendte effektivitetskontekst som optræder når vi er tvungne til fordele begrænsede ressourcer. Det er dén ene ting som teknologien ikke står til at ændre, også selvom den vil kunne lette sundhedsvæsenets byrde – i det mindst så længe vi formår at uddelegere på forsvarlig vis.

⁸ Det er værd at bemærke når IBM skriver at brugen af Watson kan lade eksperterne fokusere på det de er bedst til nemlig plejen, så synes de at antyde at Watson har en berettiget brug der går videre end blot at understøtte de kliniske beslutninger. Det er dog, som vi skal se nedenfor, et godt spørgsmål om den type kongruensstudier som nævnes i citatet overhovedet kan bruges til at begrunde en uddelegering af egentlig beslutningskompetence til Watson.

I vores behandling af Spørgsmål #1 har vi lært at der er mindst to distinktioner der er væsentlige når vi skal lave en etisk vurdering af vores case. For det første skellet mellem beslutningstagning og beslutningsunderstøttelse, som angår præcis hvilke opgaver der uddelegeres. For det andet skellet mellem strategiske og økonomisk uddelegering, som angår hvilken type af grunde der taler for at uddelegere en given opgave. Hvad mere er, vi kan nu også begynde at ane hvordan der kunne være en mere intim forbindelse mellem de to distinktioner.

Vi sagde tidligere at kræftbehandling er noget vi kerer os om og er noget som bør have stor betydning for os, og at vi derfor bør være ekstremt varsomme med at uddelegere kontrol i form af beslutningskompetencer til AI-systemer som Watson. Vi rejste også den mulige indvending at vi ikke a priori skal udelukke nogen som helst form for uddelegering til et teknologisk system, eftersom svaret på om vi bør uddelegere altid vil afhænge af i hvor høj grad systemet er i stand til at præstere bedre end de bedste menneskelige eksperter. Men hvilken form for præstationsevne er relevant her? Er det den som angår strategiske grunde eller den som angår økonomiske grunde? Den ultimative grund til at vi kerer os om hvilken kræftbehandling en patient anbefales er at vi ønsker at patienten skal modtage den bedst mulige behandling og ikke at vi ønsker at komme med anbefalinger på den mindst ressourcekrævende måde. Kunne vi f.eks. vide at Watson er signifikant bedre til at vurdere udkomme af de forskellige mulige behandlinger så ville det muligvis tale for at uddelegere mere kontrol til Watson.⁹ Dette fortæller os at hvis vi er på udkig efter grunde til uddelegere egentlig beslutningskompetence til et AI-system som Watson, så er det vi bør lede efter strategiske grunde og ikke blot økonomiske grunde.

Med de redskaber vores to distinktioner giver os, er vi nu klar til at vende os mod Spørgsmål #2, dvs. spørgsmålet om hvor god Watson er og hvordan vi kan måle dens præstationsevne.

⁹ Mange tak til Isaac Wagner for at gøre os opmærksomme på denne form for indvending og for at rejse denne specifikke udgave af den.

Spørgsmål #2: Hvordan måler vi Watsons præstationsevne? Problemet med kongruens

Som vi lige har set bliver forskellige præstationstærskler relevante alt afhængig af om man argumentere for at uddelegere opgaver til Watson ud fra strategiske eller ud fra økonomiske grunde. For at vi kan have strategiske grunde i hænde må vi have begrundet tiltro til at Watson er kvalitativt bedre end de menneskelige eksperter. De økonomiske grunde derimod kræver kun at Watson lever op til en standard sat af de relevante kræftlæger eller i det mindste nærmer sig denne standard. Argumenter for at uddelegere opgaver til en given teknologi rummer ofte en blanding de to grunde, med det er væsentligt at holde sig for øje at de to typer af grunde kan fungere uafhængigt af hinanden. Vi kan have rene strategiske grunde til at implementere en given sundhedsteknologi, nemlig i de tilfælde hvor teknologien vil øge kvaliteten af behandlingen men uden nogen ressourcebesparelse. Vi kan også have rene økonomiske grunde, nemlig i de tilfælde hvor teknologien kan frigøre dyrebare ressourcer men uden nogen kvalitetsforbedring. Det er afgørende at vi ikke blander de to typer af grunde sammen fordi vi derved risikerer at blive forledt til at tro at vi bedre grund til at uddelegere en bestemt opgave til en ny teknologi end vi egentlig har.

Men hvad ved vi egentlig om hvor god Watson er? Her er nogle resultater fra to af de tidlige forsøg på at måle Watsons præstationsniveau, hvor begge undersøger hvor stor kongruens eller overensstemmelse der er mellem Watsons anbefalinger og en specifik gruppe af kræftlægers anbefalinger:

Overall, treatment recommendations were concordant in 96.4% of lung, 81.0% of colon and 92.7% of rectal cancer cases. By tumour stage, treatment recommendations were concordant in 88.9% of localized and 97.9% of metastatic lung cancer, 85.5% of localized and 76.6% of metastatic colon cancer, and 96.8% of localized and 80.6% of metastatic rectal cancer ([Somashekhar et al 2017](#)).

The overall concordance rate was 83%; 89% for colorectal, 91% for lung, 76% for breast, and 78% for gastric cancer. Similar concordance rates were observed when

retrospective and prospective cases were analyzed separately. Discordance was attributable in part to local oncologists' preferences for non-U.S. guidelines for certain cancers, especially gastric cancer ([Suwanvecho et al 2017](#)).

Som vi så i det første IBM-citat under Spørgsmål #1, så bruger IBM overensstemmelse med menneskelige kræftlæger som en del af deres markedsføring, og det udgør også det mest udbredte kvalitetsparameter i de kendte forsøg på systematisk at teste Watson. Derfor er det væsentligt at spørge hvad vi kan lære af sådanne studier. Men før vi når til dette helt afgørende spørgsmål, er der et par ting der er værd at bemærke.

For det første skal vi ikke overse at det første af disse studier deklarerer funding fra IBM, altså firmaet der udvikler og markedsfører.¹⁰ Faktisk er der også forbindelser til IBM i det andet studie i kraft af medforfattere som er ansatte hos IBM. Men selvom det bør give anledning til en vis sund skepsis, så er en eller anden grad af samarbejde nok at forvente i disse tidlige studier, af den simple grund at studierne kræver adgang til teknologien. Vi hævder ikke at disse forbindelser til IBM diskvalificerer studierne, men det er vigtigt at være opmærksomme på det der står med småt også i denne sammenhæng.

For det andet så er de to studier udført i kontekst af andre sundhedssystemer end der hvor Watson oprindeligt blev udviklet. Mens Watson som nævnt er et resultat af et samarbejde mellem IBM og *Memorial Sloan Kettering Cancer Center* i New York, så er de to studier udført i henholdsvis Indien og Thailand. Denne diversitet rejser sine egne komplekse medicinske, økonomiske og politiske spørgsmål som vi vender tilbage til nedenfor.

For det tredje så er det værd at bemærke at det der her menes med kongruens (*concordance*) i forhold til en bestemt patient-case er at den behandling som kræftlægerne endte med at anbefale

¹⁰ De relevante oplysninger kan findes her: <https://coi.asco.org/Report/ViewAbstractCOI?id=193610>

også var at finde *enten* i blandt de behandlinger Watson kategoriserede som “Recommended” *eller* blandt de behandlinger Watson kategoriserede som “For Consideration”. Med andre ord så betyder kongruens ikke nødvendigvis et match mellem den behandling de menneskelige eksperter anbefaler og en af de behandlinger som Watson kategoriserer som anbefalede.

Efter disse tre indledende pointer så lad os kigge på de rå data. Hvad fortæller disse studier os om Watson? Er Watson tilstrækkelig pålidelige eller burde vi være nervøse for at uddelegere medicinske beslutninger eller blot beslutningsunderstøttelse til systemet? Hvad bør vi mene om en kongruens på minimum 76% og maksimalt 97,9%? Er 76% godt nok? Er 97,9% fremragende?

Som så ofte fortæller tal alene os ikke særlig meget.¹¹ Og før vi overhovedet kan begynde at besvare sådanne spørgsmål må vi først adressere et mere grundlæggende spørgsmål: Bør vi overhovedet bruge kongruens som målestok når vi skal vurdere Watsons præstationsniveau? Svaret på dette spørgsmål afhænger helt og holdent af om vi er interesserede i at afgøre om der kunne være økonomiske grunde til at implementere Watson eller om vi snarere er interesserede i strategiske grunde.

Givet at disse studier sammenligner Watsons præstation med menneskelige eksperter kunne man forledes til at tro at de kan tjene som begrundelse for en strategisk uddelegering, hvor det netop er helt afgørende hvordan en given teknologi præsterer i forhold til de menneskelige eksperter. Men det er faktisk det modsatte som er tilfældet. Antager vi at brugen af Watson kan effektivisere ressourceforbruget så synes overensstemmelse med eksperterne netop at kunne understøtte en økonomisk begrundelse for at uddelegere til Watson. Det afgørende når vi taler om økonomiske

¹¹ Det bør også nævnes at andre lokale studier viser en mindre grad af kongruens. Studier fra Sydkorea viser f.eks. en forholdsvis lav kongruensrate (~ 40%) (Choi et al. 2019; Lee et al. 2018). De Sydkoreanske hospitaler endte med ikke at implementere Watson (Choi 2018).

grunde er at den teknologi vi uddelegerer til kan løse de relevante opgaver med en kvalitet der er på højde med de menneskelige eksperter. Og en høj grad af kongruens kan netop tjene som evidens for at Watson yder på et niveau der nogenlunde svarer til de menneskelige eksperter.

Hvis vi derimod er på udkig efter strategiske grunde så synes kongruens ikke at være den rette målestok at bruge. En høj grad af kongruens kan højst vise os at systemet ikke er dårligere end de relevante menneskelige eksperter og kan derfor aldrig isoleret set fungere som evidens for at Watsons anbefalinger er bedre end dem som de menneskelige eksperter kommer op med. Hvis vi faktisk går efter et system som er mennesket overlegent så må vi per definition sigte efter et system der i en eller anden grad afviger fra den menneskelige dømmekraft. Ellers vil der ikke være nogen mulighed for at systemet rammer rigtigt der hvor mennesket tager fejl. Problemet med de aktuelle studier er i denne sammenhæng, at vi ikke ved hvad den manglende fulde kongruens skyldes. Skyldes den at maskinen rammer rigtigt i tilfælde hvor de menneskelige eksperter skyder forbi, eller måske det omvendte, eller skyldes den måske at begge parter, i hver deres retning, er lige langt fra at finde den optimale behandling til patienten? Det vi mangler, er med andre ord en standard der er uafhængig af både det svar vi får fra den relevante gruppe af menneskelige eksperter og af det output vi får fra Watson, men som ikke desto mindre udgør en standard vi har grund til at tro er mere pålidelig end både Watson og eksperterne.¹²

Kan vi få fingre i en sådan mere objektiv standard? I nogle tilfælde hvor AI-systemer bruges med diagnostiske formål, giver det god mening at tale om en sådan mere objektiv standard. Der kan vi teste pålideligheden af henholdsvis AI-systemet og de menneskelige eksperter ved at bede dem angive diagnoser på patient-cases hvor vi allerede kender det rigtige svar gennem metoder som må regnes for så godt som sikre. Men når det handler om hvilken kræftbehandling der bør anbefales, så der det ikke så let at finde en sådan 'gylden standard' (*gold standard*).

¹² Se Tupasela og Di Nucci (i review) for yderligere diskussion af dette kongruensproblem (*problem of concordance*).

Den mest pålidelige metode til at bestemme det sandsynlige udfald af en given kræftbehandling til en given patient er formentlig at bruge resultaterne fra prospektive randomiserede kliniske forsøg med patientgrupper der sammenlignelig med den pågældende patient. Men i mange faktiske tilfælde besidder vi ikke en tilstrækkelig evidensbasis for at kunne foretage en præcis vurdering af det sandsynlige udfald af en given kræftbehandling for en konkret patient. Dette er en af grundene til at vi finder uenighed mellem eksperter når de skal vurdere hvilken behandling der er mest hensigtsmæssig for en konkret patient.¹³ Det er samtidig en af grundene til at IBM bruger input fra eksperterne fra MSK *Cancer Center* når de træner Watson.¹⁴

Selv hvis vi havde flere kliniske forsøg til rådighed ville anvendelsen af de generelle resultater på et partikulært tilfælde stadig involvere et uudryddeligt element af klinisk dømmekraft på en måde som f.eks. testmetoder der bruges som gyldne standarder i diagnostisk sammenhæng ikke gør det. Dette bør imidlertid ikke få os til at udelukke at AI-systemer som Watson kunne blive bedre end menneskelige eksperter til at forudsige en given behandlings udkomme i et konkret tilfælde. Hvis Watson, som IBM mener den vil, bliver i stand til at opfange relevante og ellers skjulte mønstre i de enorme datamængder den processerer, så kan denne evne måske ende med at gøre den de menneskelige eksperter overlegne. Men vores erkendelsesteoretiske spørgsmål var hvordan vi nogensinde kan vide at maskinen overvinder den menneskelige dømmekraft. I fraværet af en gylden standard vil den bedste indikation på at en sådan overlegenhed måske være at de menneskelige eksperter, efter modne overvejelser af den evidens Watson præsenterer dem for, er villige til i et betydeligt antal tilfælde at indrømme at Watsons anbefalinger bør følges.

¹³ Somashekhar et al 2018 skriver følgende i det kongruensstudie vi citerede ovenfor: "Finally, lack of concordance does not necessarily provide evidence regarding whether the MMDT or WFO was 'correct' in its recommendation. As previously discussed, disagreements can have many valid explanations such as differences in how patients with comorbidities and increasing age are treated. There is no 'gold standard' beyond expert opinion, which studies have shown can vary profoundly during the evaluation of CDSSs."

¹⁴ Ross og Swetlitz (2017) beskriver i deres kritiske nyhedsartikel disse træningssessioner med Watson: <https://www.statnews.com/2017/09/05/watson-ibm-cancer/>.

Ret skal være ret, det forhold at kongruens-studier ikke kan tjene som strategiske grunde for at implementere Watson er ikke det samme som at der ikke er strategiske grunde der taler for at bruge Watson. Selv hvis vi antager at Watsons anbefalinger ikke er kvalitativt bedre end kræftlægenes så kunne Watson stadigvæk godt være med til at kvalificere lægenes beslutninger. Eftersom Watson kan skanne massive mængder af litteratur på få sekunder så er det forventeligt at den kan identificere evidens som klinikerer ikke er bekendt med, og som kan bekræfte eller udfordre dennes egen vurdering. Her er det væsentligt at det netop er en specifik opgave i relation til at understøtte klinikerens beslutning som Watson muligvis løser bedre end klinikerer selv og altså ikke selve den opgave der består i at fremkomme med de bedst mulige anbefalinger. Dette er væsentligt fordi det betyder at vi stadigvæk er langt fra at have fundet en strategisk grund til at uddelegere egentlig beslutningskompetence i forhold til hvilken behandling en patient bør tilbydes. Og for så vidt det er korrekt, som vi argumenterede for under Spørgsmål #1, at vi har brug for stærke strategiske grunde hvis vi skal så meget som overveje at uddelegere kontrol over så livsvigtige beslutninger til et system som Watson, så kan vi konkludere at vi ikke besidder grunde til at overveje en sådan uddelegering på nuværende tidspunkt.

Alle disse komplekse erkendelsesteoretiske vanskeligheder bør ikke får os til at miste blikket for en langt mere enkel iagttagelse: Hvis Watson leverer anbefalinger som fælleskabet af videnskabelige eksperter er enige om er forkerte og måske direkte skadelige, så betyder det at dette specifikke medicinske AI-system endnu ikke er klar til brug. Og der er noget der tyder på Watson at Watson stadig har problemer i denne henseende. I 2018 var der medierapporter om at Watsons ifølge IBMs egne interne dokumenter i flere tilfælde kom med anbefalinger der var 'unsafe and incorrect'.¹⁵

¹⁵ Se for eksempel de følgende tre, hvor den første synes at være kilden til de øvrige historier:

Spørgsmål #3: Yderligere risici

Indtil videre har vi haft fokus på de erkendelsesteoretiske problemer der rejser sig når vi skal vurdere et medicinsk AI-system som Watson. Men ud over de usikkerheder og vanskeligheder som dukkede op i den forbindelse, hvilke andre risici kan vi så identificere i forbindelse med sundhedsrelaterede AI-systemer som Watson? Et problem er det såkaldte ‘black box’-problem (Pasquale 2015) og de beslægtede bekymringer omkring systemernes *opacitet* eller uigennemsigtighed. Kogt helt ned er bekymringen at vi simpelthen ikke forstår hvordan algoritmerne fungerer. Vi kan også forklare bekymringen ved at fremhæve det åbenlyse, nemlig at opacitet eller uigennemskuelighed er det modsatte af transparens eller gennemskuelighed. Og transparens er specielt afgørende i en sundhedsfaglig sammenhæng, bl.a. som en nødvendig betingelse for informeret samtykke (se Di Nucci 2018, Kapitel 6). Derfor bør vi bør kigge lidt nærmere på opacitets-fænomenet.

Burrell (2016) udpeger tre forskellig slags opaciteter som kan optræde i forbindelse med maskinlæringsalgoritmer:

- (i) Opacitet via ophavsrettigheder og forretningshemmeligheder;
- (ii) Opacitet pga. teknisk analfabetisme (f.eks. manglende evne til at læse kodesproget);
- (iii) Opacitet som skyldes “mismatch between mathematical procedures of machine learning algorithms and human styles of semantic interpretation” (Burrell 2016: 3).

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

<https://www.telegraph.co.uk/technology/2018/07/27/ibm-watson-ai-criticised-giving-unsafe-cancer-treatment-advice/>

<https://www.dailymail.co.uk/sciencetech/article-6001141/IBMs-Watson-suggested-inaccurate-unsafe-treatment-recommendations-cancer-patients.html>

<https://gizmodo.com/ibm-watson-reportedly-recommended-cancer-treatments-tha-1827868882>

De tre former for opacitet henviser til tre forskellige måder vi kan efterlades i et mørke når vi prøver at forstå hvordan et AI-system løser sin opgave. Opacitet via ophavsrettigheder og forretningshemmeligheder er en ægte bekymring i tilfælde som IBMs Watson, hvor et firmas økonomiske interesser kan være i åbenlys konflikt med offentlighedens interesse i at opnå det bedst mulige grundlag for at teste systemet videnskabeligt. Opacitet pga. teknisk analfabetisme er også relevant i vores case, eftersom læger som flest vil mangle den tekniske indsigt det kræver at forstå Watsons algoritmer. Man kan mene at dette er et naturligt og uskyldigt resultat af en fornuftig arbejdsdeling, men fordi vi ikke besidder en simpel objektiv ‘gylden standard’ i henhold til hvilken vi kan vurdere Watsons output, bliver det afgørende at lægerne har mulighed for at vurdere om Watsons konklusioner er velbegrandede. Hvis ikke lægerne kan anerkende at Watsons konklusioner hviler på gode grunde, vil deres eventuelle tillid til Watson være blind. Den tredje form for opacitet har at gøre med netop den principielle mulighed for at oversætte algoritmernes “rationalitet” til ræsonnementer der kan følges af mennesker. Bekymringen er at systemernes evne til mønstergenkendelse kan blive så kompleks, at det ikke længere blot er manglende teknisk viden der står i vejen for at begribe hvad maskinerne foretager sig. Tanken er at der kan opstå en egentlig inkommensurabilitet eller uoversættelighed mellem måden systemets algoritmer virker og den måde mennesket tænker i naturlige sprog. Denne bekymring anser vi for at være af stor både praktisk og teoretisk betydning, men det vil lede os for langt væk fra vores egentlige formål at forfølge den yderligere her.

Helt generelt kan vi sige at en høj grad af epistemisk transparens, som kan tillade at menneskelige eksperter kan tjekke om Watsons anbefalinger er velbegrandede, er fuldstændig afgørende fordi vi ikke har en uafhængig objektiv standard der kunne tillade os at skyde genvej til det rigtige resultat uden at skulle omkring menneskelige dømmekraft. Så må det vise sig hen af vejen om denne

mulighed for at gå Watson efter i sømmene svinder ind, og det måske af principielle grunde, efterhånden som Watsons evne til mønstergenkendelse øges.

Watson løser ikke blot en ren teoretisk opgave, som f.eks. den teoretiske opgave at nå frem til hvad den mest plausible diagnose for en given patient er. Watson kommer også med praktiske anbefalinger om hvilken behandling en patient bør tilbydes. Det har den konsekvens at det ikke blot er Watsons epistemiske standarder, f.eks. hvordan Watson vægter forskellige typer af evidens, som bør være transparente, men også de værdier som ligger bag Watson anbefalinger. Dette værdielement kommer med egne risici som går ud over de ovenstående opacitetsbekymringer. McDougall (2019) gør f.eks. opmærksom på at Watsons primære kriterium er 'disease-free survival'. Ifølge McDougall opstiller Watson en prioriteret rækkefølge af mulige behandlinger ud fra sandsynligheder for sygdomsfri overlevelse, beregnet på grundlag af eksisterende evidens i den medicinske litteratur som Watsons algoritmer processerer.¹⁶

Det er nyttigt her at holde fokus på 'sygdomsfri overlevelse' fordi selv så tilsyneladende ukontroversielt et kriterium rejser grundlæggende spørgsmål: Bør sygdomsfri overlevelse altid være det primære behandlingsmål uanset hvad omkostninger er? Erfaringerne fra palliativ pleje og i det hele taget med patienter der skal træffe beslutninger sidst i livet viser med al tydelighed at statistikker om overlevelseschancer ikke altid er det mest relevante for patienten.

Problemet her er ikke blot at vi bruger ét bestemt kriterium eller hvordan vi begrundet valget af et kriterium frem for et andet. Sådanne valg er omdiskuterede og vanskelige uanset om algoritmer er involverede eller ej. Det nye problem vi får når algoritmer kommer på banen er at det ændrer den dialektiske situation som sådanne kontroversielle variabler som 'sygdomsfri overlevelse' indgår i. Så snart IBM vælger 'sygdomsfri overlevelse' som Watsons hovedhensyn så vil selve det valg i sig

¹⁶ Det er uklart for os præcis hvordan Watson vægter de input om konkrete cases den modtager gennem træningssessionerne med eksperterne på MSK Cancer Center og den evidens algoritmerne identificerer på egen hånd. Se Ross & Swetlitzs (2017) artikel for en udlægning af hvordan Watson virker her.

selv skubbe til den magtdynamik der omgiver beslutningstagning i forbindelse med kræftbehandling. Dette vil ske fordi en bestemt mulighed, som tidligere var kontroversiel, nu er blevet valgt som den man skal tage udgangspunkt i (denne bekymring kan sammenlignes med en bekymring om at smartphone- og app-udviklere foretager designvalg der gør en bestemt (måske ikke-ideel) opsætning til *default mode* med store (negative) konsekvenser for hvordan mennesker interagerer med teknologien).

Vi kan også udtrykke denne pointe således: Tal har allerede den fordel (specielt i et teknologisk samfund) at vi kan operationalisere dem. Denne position bliver bestyrket når magtfulde spillere vælger variabler der kan kvantificeres frem for variabler der vanskeligt kan kvantificeres. Når valget falder på sådanne kvantificerbare variabler så er det åbenlyst netop fordi de er nemme at operationalisere. Men her må vi passe på ikke at forudsætte det som burde være spørgsmålet nemlig om kvantificerbare variabler altid er de rigtige at vælge. Den afgørende grund til at vælge en variabel (i vores tilfælde 'sygdomsfri overlevelse') frem for en anden, kan ikke alene være muligheden for at operationalisere denne variabel. I så fald risikerer vi at lukke ned for den livlige debat om hvordan f.eks. beslutninger der angår døende patienters skal træffes. Teknologien ville ende med at retfærdigøre sig selv, hvis det antages at en variabel kun er værdifuld hvis den kan operationaliseres og altså er teknologivenlig. Udtrykt i et slogan: Værdierne må have værdi ud over deres værdi for teknologien.

Der er andre risici forbundet med Watson der er uafhængige af vores evne til at måle Watsons præstationsevner. En særlig interessant risiko vi ønsker at fremhæve angår det globale sigte som Watsons har. Som allerede nævnt så er de tidlige kongruensstudier af Watson hovedsageligt udført i anden- og tredje verdens lande. På den ene side er dette gode nyheder. Hvis vi via software kan bringe onkologisk ekspertise ud til lande og sundhedssystemer som ikke selv har ressourcer til at udvikle en sådan ekspertise, så er der håb om at Watson kan bidrage til at løse globale

sundhedsudfordringer og til at opnå større global retfærdighed. Det er noget vi bør byde velkommen.

Men på den anden side, er der også en række vigtige forbehold der bør fremhæves. For det første: Hvis disse sundhedsvæsener mangler midler til selv at udvikle den nødvendige ekspertise, så er der fare for at købet af licens til Watson blot vil forværre den økonomiske situation. Muligvis vil de sparsomme ressourcer i f.eks. et land som Indien være bedre brugt på at investere i sundhedsuddannelse frem for teknologier fra andre lande, i det mindste hvis vi tænker på muligheden for en langsigtet bæredygtig udvikling.

For det andet: Hvis disse lande og sundhedssystemer ikke på forhånd har den nødvendige onkologiske ekspertise, på hvilket grundlag skal de så kunne vurdere Watsons præstationsevne? Det er klart at kongruens ikke vil være den rigtige målestok her, eftersom hele pointen med at bruge Watson ville være at forbedre den aktuelle standard. Med andre ord der det vi er på udkig efter strategiske og ikke økonomiske grunde til at implementere Watson.¹⁷ Der er en reel fare for en form for medicinsk kolonialisme hvor der opstår et urimeligt stort pres på lokale læger for at de skal acceptere udefra kommende ekspertise. Et problem her er at hvis Watsons output i høj grad er et resultat af træningssessionerne med MSK-eksperterne i New York, så kan vi ikke forvente at Watson tager højde for hvordan kulturelle (inklusive f.eks. spisevaner) og miljømæssige faktorer kan spille ind i forhold til hvilken behandling der er anbefalelsesværdig.

Vi vil slutte behandlingen af Spørgsmål #3 af ved at tilføje en fjerde risikokategori til de tre ovenstående, dvs. til de transparensrelaterede risici, de værdirelaterede risici og de risici som knytter sig til globaliseringspolitik indenfor sundhedssektoren: Risikoen for at læger, det være sig i

¹⁷ Implementering på baggrund af manglende lokal ekspertise finder allerede sted. Se Ross og Swetlitzs artikel (2017) for en beskrivelse af en sådan implementering på et hospital i Mongoliet.

New York, Indien eller København, kommer til at fæste for meget lid til Watson.¹⁸ Dette fænomen er velkendt indenfor luftfart, hvor piloter ofte bliver bedt om at lande et fly manuelt selvom automatpiloten er fuldt ud i stand til det, for simpelthen at sørge for at piloterne ikke glemmer hvordan man gør og holder deres færdigheder ved lige.¹⁹

Hvordan sikrer vi at kræftlæger ikke bliver for afhængige af Watson? Et svar på det spørgsmål kræver uden tvivl at vi tager fat på det tæt forbundne spørgsmål om hvad det vil sige at bruge Watson for meget. I brugen af Watson som beslutningsunderstøttende som vi har beskrevet den her, kan alle være enige om at det ville være at bruge Watson for meget, hvis en læge f.eks. ukritisk videregav Watsons anbefalinger til patienten uden selv først at vurdere anbefalingernes plausibilitet. Derved ville lægen f.eks. være skyld i at de mulige *algoritmiske fordomme* ('algorithmic biases') i forhold til behandlingsvalg som kan være indeholdt i programmet bliver videreført og bekræftede.²⁰

Der er andre mulige scenarier der ville udgøre eksempler på overdreven brug af Watson, f.eks. hvis kræftlægen på et spørgsmål om gavnligheden af en given behandlingsform svarer "Det var Watsons førstevalg" eller "Det er hvad Watson anbefaler". Dette ville være lige så slemt og faktisk værre end hvis lægen svarede "Det er mit førstevalg" eller "Det er hvad jeg anbefaler". Ingen af disse svar lever op til de grundlæggende krav til god kommunikation og til informeret samtykke, men det sidste svar tillader, i det mindste i teorien, at patienten kan udfordre den kilde som fremkommer med anbefalingen, nemlig lægen som de taler med. I den første type af svar vil Watson derimod figurere som en fraværende autoritet i form af skjulte og for patienten formentlig uigennemskuelige algoritmer.

¹⁸ Faktisk har Rigshospitalet i København kørt et pilotforsøg med Watson, Se Tupasela & Di Nucci (i review) for flere detaljer om dette forsøg. Bl.a. på baggrund af lave kongruenstal valgte Rigshospitalet ikke at købe licens til Watson.

¹⁹ <https://eu.usatoday.com/story/travel/columnist/cox/2014/02/09/autoland-low-visibility-landings/5283931/>

²⁰ Der findes et væld af nyere forskningsartikler der peger på problemer med algoritmiske fordomme, se f.eks Hajian et al. 2016, Bozdag 2013, Baeza-Yates 2016, Kirkpatrick 2016, Lambrecht & Tucker 2018, og Danks & London 2017.

Der findes imidlertid også langt mere subtile måder hvorpå lægerne kan ende med at fæste for meget lid til Watson. Over tid kunne der ske en tilvænning til Watson og til de svar som algoritmerne kommer med, således at de menneskelige eksperter tiltro til Watsons anbefalinger langsomt vokser og der til sidst ikke vil blive stillet nok kritiske spørgsmål til dem. Dette er et specielt skræmmende scenarie da det vil kunne betyde at algoritmiske fordomme vil lede til menneskelige fordomme og vi vil få en ond cirkel af gensidigt bekræftende fordomme og fejlbedømmelser.

Selvom det vil være vanskeligt og kræve langt mere begrebsligt arbejde præcist at definere hvad det vil sige at fæste for meget lid til et AI-system som Watson, så kan alle nok være enige om at der er scenarier, som de ovenfor beskrevne, der udgør eksempler på en sådan overdreven tiltro. Derfor vil det også være vigtigt at etablere processer der sikrer at læger ikke ender med at tildele Watson for meget autoritet og at de bestandigt træner og udvikler deres evne til kritisk at bedømme Watsons anbefalinger.

Spørgsmål #4 og konkluderende bemærkninger

Afslutningsvis vil vi vende tilbage til det grundlæggende etiske spørgsmål om hvad vi på ene side kan tillade os at uddelegere til Watson og hvad vi på den anden side bør afholde os fra at bruge Watson til. Vi har skelnet mellem at uddelegere selve beslutningskompetencen til Watson og så det at uddelegere beslutningsunderstøttelse (rådgivning til klinikerens); og vi har anbefalet, i det mindste i denne tidlige fase, at kun den understøttende opgave (om nogen overhovedet) bør tildeles Watson.

Desuden har vi skelnet mellem såkaldt strategisk uddelegering, hvor målet er en kvalitativ forbedring af præstationsniveauet, og så økonomisk uddelegering, hvor sigtet er at bevare præstationsniveauet og samtidig bruge mindre ressourcer. Vi identificerede nogle mulige problemer og risici ved strategisk uddelegering til Watson, inklusiv det forhold at kongruensstudierne ikke

understøtter strategiske grunde til uddelegering og at det er afgørende at bibeholde menneskelig kontrol over processer som er så vigtige for os som kræftbehandling er. De økonomiske grunde til at uddelegere til Watson synes derimod mere lovende, i hvert fald så længe vi holder os til at uddelegere basale opgaver som at gennemtrawle den eksisterende litteratur og derved frigøre noget af kræftlægenes kostbare tid.

Derudover har vi identificeret nogle mulige problemer og risici som er uafhængige af præcis hvor god Watson er og af problemet med at måle dens præstationsevne. Vi identificerede i alt fire typer af bekymringer: 1) de transparensrelaterede; 2) de værdirelaterede, særligt i forbindelse med beslutninger ved livets afslutning; 3) bekymringer i forhold global sundhedsretfærdighed; 4) og endelig risikoen for at vi i sundhedssystemet bliver for afhængige af Watson og den dermed forbundne fare for fremme af algoritmiske fordomme og langsom undergravning af de menneskelige eksperters dømmekraft.

Vi håber at vores forsøg på i hvert fald at lægge grundstenene til en grundig etisk vurdering af Watson kan være brugbar som case på mindst tre måder: For det første som et eksempel på hvordan man kan gå til værks når man skal vurdere om et konkret informationsteknologisk system kan benyttes sikkert og hensigtsmæssigt i sundhedsvæsenet; for det andet som et eksempel der illustrerer hvordan etiske spørgsmål ofte er sammenfiltrede med komplekse filosofiske spørgsmål, som f.eks. de erkendelsesteoretiske spørgsmål vi har diskuteret i forbindelse med kongruensstudierne; endelig håber vi at analysen tjener som et praktisk eksempel på hvordan filosofi og etik faktisk kan *bruges* i den virkelige verden.²¹

²¹ Vi retter en stor tak til Isaac Wagner for hans skriftlige kommentarer til en tidligere version af denne artikel.

Referencer

- Baeza-Yates, R. 2016, May. Data and algorithmic bias in the web. In *Proceedings of the 8th ACM Conference on Web Science* (pp. 1-1). ACM.
- Bozdag, E. 2013. Bias in algorithmic filtering and personalization. *Ethics and information technology*, 15(3), 209-227.
- Burrell J. 2016. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1) :1-12.
- Youn I Choi, Jun-won Chung, Kyoung Oh Kim, et al. 2019. Concordance Rate between Clinicians and Watson for Oncology among Patients with Advanced Gastric Cancer: Early, Real-World Experience in Korea. *Canadian Journal of Gastroenterology and Hepatology*, 2019, Article ID 8072928. <https://doi.org/10.1155/2019/8072928>.
- Choi, M. 2018. “Major Hospitals in S. Korea not very interested in Watson.” 2 March 28, 2018. <http://www.businesskorea.co.kr/news/articleView.html?idxno=21308>. Visited May 14.
- Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (pp. 4691-4697).
- Di Nucci E. 2018. *Ethics in Healthcare*. Rowman & Littlefield (danish edition : *Ethics4Medics*, Munksgaard 2018).
- Di Nucci E. 2019. Should we be afraid of medical AI? *Journal of Medical Ethics* 45(8): 556-558.
- Domingos P. 2015 *The Master Algorithm*. Basic Books.
- Fry H. 2018. *Hello World*. Penguin.
- Grote & Di Nucci (forthcoming). Algorithmic Decision-Making and the Problem of Control. In: Beck B & Kühler M (eds) *Anthropology, Technology and Responsibility*. Metzler.
- Hajian, S., Bonchi, F., & Castillo, C. 2016. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 2125-2126). ACM.
- Kirkpatrick, K. 2016. Battling algorithmic bias: How do we ensure algorithms treat us fairly? *Communications of the ACM*, 59(10), 16-17.
- Lambrecht, A., & Tucker, C. E. 2018. Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM career ads. *Gender Action Portal*:

<http://gap.hks.harvard.edu/algorithmic-bias-empirical-study-apparent-gender-based-discrimination-display-stem-career-ads>

Lee et al. 2018. Assessing Concordance With Watson for Oncology, a Cognitive Computing Decision Support System for Colon Cancer Treatment in Korea. *JCO Clinical Cancer Informatics*, 2(2): 1-8.

McDougall R. 2019. Computer knows best? The need for value-flexibility in medical AI. *Journal of Medical Ethics* 45 (3): 156.

Mittelstadt, B. D., & Floridi, L. (Eds.). (2016). *The ethics of biomedical big data* (Vol. 29). Springer.

Pasquale F 2015. *The Black Box Society*. Harvard UP.

Ross, C & Swetlitz I. 2017 (5. September). IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. *STAT*: <https://www.statnews.com/2017/09/05/watson-ibm-cancer/>

Santoni, F. & Van den Hoven, J. 2018. Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, 5, 15.

Somashekhar et al 2017. *J Clin Oncol* 35, 2017 (suppl; abstr 8527)

Susskind R & Susskind D (2015), *The Future of the Professions*. Oxford UP.

Suwanvecho et al 2017. *J Clin Oncol* 35, 2017 (suppl; abstr 6589)

Thaddeus Beck et al 2017. *J Clin Oncol* 35, 2017 (suppl; abstr 6501)

Tegmark M. 2017. *Life 3.0*. Penguin.

Topol E. 2019. *Deep Medicine*. Basic Books.

Tupasela A & Di Nucci E. (submitted). The Problem of Concordance as Evidence in Oncology Decision-Support Systems. *AI & Society*.

Wajcman J 2017 Automation : is it really different this time ? British Journal of Sociology 68 (1) : 119-127

STUDIESPØRGSMÅL

- I) Hvad er forskellen mellem beslutningstagning og beslutningsunderstøttelse? Og mere specifikt: Hvad er forskellen på at uddelegere beslutningskompetence til et teknologisk system og så at uddelegere beslutningsunderstøttende opgaver til et sådant system.
- II) Hvad er forskellen mellem strategisk og økonomisk uddelegering?
- III) Angiv mindst tre potentielle risici ved Watson og vurder hvor sandsynlige og plausible disse tre risici er.
- IV) Diskuter, med udgangspunkt i artiklens analyse af Watson, om vi er ved at miste det menneskelige aspekt i sundhedsvæsenet.
- V) Sammenlign Watson-casen med andre aktuelle sundhedsteknologier.