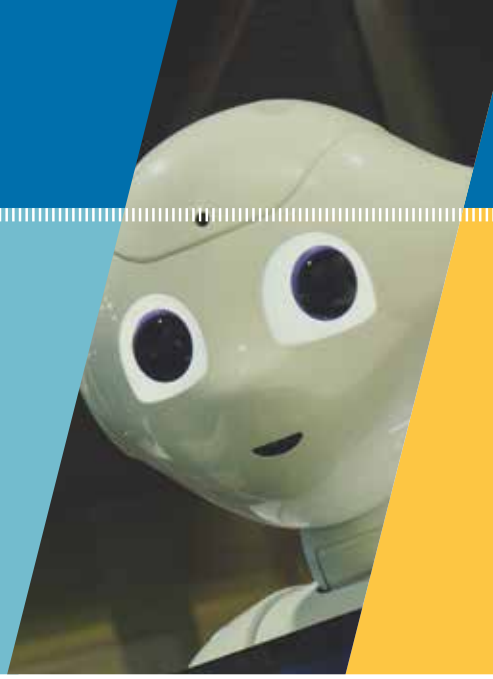
STATENS MEDICINSK-ETISKA RÅD



Kort om

Artificiell intelligens i hälso- och sjukvården

Det finns stora förväntningar om att artificiell intelligens (AI) ska bidra till att lösa framtidens utmaningar inom hälso- och sjukvården. Samtidigt medför AI många etiska och samhälleliga utmaningar. AI har klassats som en av de framväxande teknologier som har störst nyttopotential, men också medför högst risk för negativa konsekvenser.[1]

Vad är AI?

Det finns ingen allmänt etablerad definition av AI. John McCarthy m.fl., som 1955 myntade begreppet artificiell intelligens, definierade AI som *vetenskapen och tekniken att skapa intelligenta maskiner*.[2] Numera beskrivs AI ofta som datorsystem som kan utföra uppgifter som tidigare ansetts kräva mänsklig intelligens, såsom igenkänning av tal och bilder, översättning av språk, inlärning, problemlösning och beslutsfattande. Även forskningsområdet som syftar till att utveckla sådana system kallas för AI.

De senaste åren har utvecklingen inom AI accelererat, och det rapporteras dagligen om nya tillämpningar inom snart sagt alla områden i samhället. Viktiga faktorer bakom de framsteg som skett är kraftfullare datorprocessorer, den snabbt växande tillgången på data (Big Data) samt utvecklingen av allt effektivare algoritmer.

De flesta i dag kända tillämpningar av AI är av den typ som kallas för *snäv AI*. Här är syftet att automatisera en specifik uppgift som i dag utförs av människor, oftast med målet att öka snabbheten, skalbarheten och/eller tillförlitligheten. Detta kan kontrasteras mot det som kallas *generell AI*, där syftet är att skapa datorprogram som i likhet med människor förstår och resonerar på en mer generell nivå. Denna skrift behandlar i huvudsak tillämpningar av snäv AI.

Andra begrepp som förekommer i samband med AI är *maskininlärning* och *djup maskininlärning* (djupinlärning). Maskininlärning använder algoritmer som möjliggör för ett AI-system att ”lära” sig att utföra en uppgift utifrån data utan att ha getts specifika instruktioner för just den uppgiften. Djup maskininlärning är en särskild form av maskininlärning som använder algoritmer som inspi-

Publikationsserien ”Kort om” lyfter fram etiska och samhälleliga aspekter inom ett aktuellt ämne.

Publicerad mars 2020

rerats av hur den mänskliga hjärnan antas bearbeta intryck och skapa mönster för beslutsfattande.

Ytterligare ett begrepp är *augmented intelligence*. Här ligger fokus på tillämpningar som stöder och förbättrar den mänskliga intellektuella förmågan, snarare än ersätter den. Människa och maskin blir ”kollegor” och arbetar sida vid sida. Just inom hälso- och sjukvården har det förutspåtts att ”förhöjning” av mänskliga förmågor genom AI är en mer sannolik utveckling än ren automatisering, åtminstone i ett första skede.[3]

Tre typer av maskininlärning

Vid maskininlärning ”lär” sig algoritmer att lösa sin uppgift genom att ”tränas” på stora mängder data. Vid *övervakad maskininlärning* innehåller träningsexemplen både indata (exempelvis pixelvärden i en bild) och utfallsvärden (vad bilden föreställer). Algoritmen upptäcker mönster i indata som korrelerar med utfallsvärdena och skapar modeller för att förutsäga utfallet när den presenteras för nya indata.

Vid *oövervakad maskininlärning* innehåller träningsdata endast indata och algoritmen sorterar själv upp exemplen utifrån mönster i indata. Vid oövervakad maskininlärning kan nya, okända samband upptäckas.

Vid *förstärkningsinlärning* lär sig en algoritm att lösa en uppgift genom att pröva olika lösningar och få återkoppling på hur väl den lyckats.

Ett AI-system kan vara statiskt, vilket innebär att modellen inte ändras när träningsfasen är över. Men det kan också vara dynamiskt och kontinuerligt justera modellen utifrån nya data som det stöter på.

AI i hälso- och sjukvård

I takt med att AI-algoritmer blir allt bättre på att sortera data, hitta mönster och göra förutsägelser, ökar förhoppningarna om att AI kan få en avgörande betydelse för att göra den framtida hälso- och sjukvården mer effektiv och säker, och samtidigt bidra till att lösa de utmaningar vi står inför med en åldrande befolkning, begränsade resurser och svårigheter att rekrytera kvalificerade medarbetare. Forskning och utveckling kring tillämpningar av AI i hälso- och sjukvården pågår i dag inom en lång rad områden, från diagnostik och behandling till administration och medicinsk forskning. Fortfarande är det dock relativt få system som bygger på maskininlärning som används i praktisk klinisk verksamhet.[4]

Bedömningar och beslut

Stora förhoppningar finns om att AI ska leda till bättre medicinska bedömningar till lägre kostnader. Genom att AI-system bygger sina förutsägelser på stora mängder patientfall är förhoppningen att bedömningarna ska kunna bli mer träffsäkra.[5] Omotiverade skillnader i praxis mellan enskilda vårdanställda skulle kunna reduceras liksom risken för fel som beror på kognitiva misstag (information som förbises, felslut eller bias).[6] AI skulle också kunna hjälpa vården att hålla riktlinjer uppdaterade genom att söka efter relevant litteratur när mängden studier på medicinområdet växer allt snabbare.[7]

Diagnostik

Framförallt inom bilddiagnostiken har en lång rad tillämpningar av AI utvecklats de senaste åren. AI har använts för att exempelvis klassificera hudcancer, hitta metastaser vid cancersjukdom, identifiera embryon för provrörsbefruktning, upptäcka livmoderhalscancer, bedöma lungröntgenbilder samt tidigt detektera förändringar i näthinnan.[8] Algoritmer som bygger på textanalys har använts för att diagnosticera barnsjukdomar[9]. Forskare har också visat att AI-baserade system kan upptäcka depression utifrån text- och talinformation.[10] I många av fallen har AI presterat lika bra som eller bättre än läkare och betydligt snabbare. Inom bilddiagnostiken förutspås AI ta över allt mer redan inom de närmaste

”Stora förhoppningar finns om att AI ska leda till bättre medicinska bedömningar till lägre kostnader.”

åren, medan det kliniska genomslaget inom allmän diagnostik kan ta längre tid.[11]

Prognostik och prevention

AI-algoritmer har rapporterats kunna förutsäga risk för sepsis (blodförgiftning) och hjärtattack med högre precision än etablerade skattningsinstrument.[12] Algoritmer har också utvecklats som letar brett efter tecken på sjukdom eller risk för sjukdom utifrån den samlade data som finns i patientjournalen: diagnoskoder, förskrivningar, labbsvar, behandlingar, fritext m.m.[13] I en studie lyckades en algoritm utifrån journaldata med hög precision förutsäga inskrivningstid, mortalitet och risk för återinläggning bland sjukhuspatienter.[14] AI-baserade journalgranskningar skulle kunna användas för förebyggande insatser, exempelvis genom att larma om att patienten behöver få sin läkemedelsbehandling justerad eller bli kallad till undersökning.

Beslutsstödssystem

AI:s roll i vården behöver inte begränsas till att ställa diagnos och göra prognoser. Många ser även framför sig att AI i framtiden kommer att spela en viktig roll som stöd för medicinskt beslutsfattande, exempelvis genom att utifrån diagnos och annan patientdata ge rekommendationer om behandling. Några områden där lovande resultat rapporterats är bröstcancer och sepsis.[15] Algoritmer har också utvecklats för att personanpassa strålbehandling i syfte att minska skador på omgivande vävnad.[16] Det har också föreslagits att AI skulle kunna stödja medicinskt beslutsfattande genom att, baserat på journalinformation men även information från exempelvis sociala medier, förutsäga en patients preferenser vid nedsatt beslutsförmåga.[17]

Ett område där AI förutspås få stor betydelse är så kallad precisionsmedicin, som handlar om att skräddarsy den mest optimala behandlingen för varje patient utifrån genetiska faktorer, sjukdomshistoria, livsstil, omgivningsfaktorer m.m. Givet de stora datamängder som behöver analyseras har det hävdats att precisionsmedicin inte kommer att vara möjlig utan AI.[18]

Andra typer av beslut inom vården där många ser en roll för AI är prioriteringsbeslut, det vill säga vilken patient som bör få behandling om resurserna inte räcker för alla, beslut om vilken patient som bör få ett donerat organ och triageringsbeslut, det vill säga när man avgör vilken patient som ska få behandling först och/eller på vilken vårdnivå.[19]

Praktiskt vårdarbete

Det är inom bedömningar och beslutsfattande som utvecklingen av AI-system inom vården kommit längst. Men det diskuteras allt mer hur AI-teknologi, ofta kopplat till robotik, även ska kunna utnyttjas i det praktiska vårdarbetet. Exempel är AI-stödda operationsrobotar, robotar som övervakar patienter på akutmottagningen, eller applikationer som understödjer telemedicin.[20] Målet kan vara att öka kvaliteten, kompensera för personalbrist eller, som vid telemedicin, överbrygga fysiska avstånd.

Proteser m.m.

En rad olika projekt pågår där AI används för att öka precisionen och funktionaliteten i avancerade proteser.[21] Försök har även gjorts att använda AI för att styra förlamade kroppsdelar med tanken. I en studie fick en patient elektroder som läser av hjärnans aktivitet inopererade i hjärnan. En algoritm lärde sig översätta denna aktivitet till elektriska signaler som stimulerar underarmens muskler.[22]

Patientinriktade tillämpningar

Allt fler patientinriktade digitala applikationer som hälsoappar och/

> "Givet de stora datamängder som behöver analyseras har det hävdats att precisionsmedicin inte kommer att vara möjlig utan AI."

eller kroppsnära teknik (wearables och internables) använder AI för att förbättra funktionaliteten, exempelvis genom att ge mer individanpassade rekommendationer. AI-baserade hälsoappar har utvecklats som vid misstänkt sjukdom bedömer symtom och ger förslag på diagnos eller hjälper patienten att hitta rätt vårdnivå.[23] Kommunikationen med patienterna kan ske via exempelvis chatbottar eller datorgenerade ”virtuella hälsoassistenter”.[24] Andra tillämpningar har som syfte att följa och stödja patienterna i deras egenvård, exempelvis genom att hjälpa patienter att optimera sin behandling eller att ta sina läkemedel i rätt tid och i rätt dos.[25] Kopplade till medicinsk utrustning för hemmabruk skulle AI-baserade appar kunna användas för att övervaka patientens värden och slå larm vid avvikelser, vilket skulle kunna leda till att sjukhusinläggningar undviks.

Administration, flöden och logistik

Vid Johns Hopkinssjukhuset i USA används AI för att fördela vårdplatser, vilket lett till att vårdplatsutnyttjandet effektiviserats.[26] Andra exempel på tillämpningar av AI på vårdadministrativ nivå är system för att lägga schema, upptäcka bedrägerier (i försäkringsfinansierade sjukvårdssystem), lösa organisatoriska utmaningar vid telemedicin samt automatisera administrativa sysslor i vården.[27]

Forskning och utveckling

Stora förhoppningar knyts till att AI ska kunna effektivisera utvecklingen av nya läkemedel. Algoritmer har utvecklats som utifrån stora mängder data ska lära sig hur molekyler interagerar och därmed kunna ge förslag på potentiella läkemedelskandidater.[28] I Storbritannien har en databas byggts upp som kombinerar genetiska och kliniska patientdata med multidisciplinär vetenskaplig information och använder AI för att generera hypoteser för läkemedelsutveckling.[29] Andra forskare har utvecklat en ”forskande robot” som med hjälp av AI ska göra processen med att upptäcka nya läkemedel snabbare och billigare.[30]

Ytterligare tänkbara tillämpningar inom medicinsk forskning kan vara att hitta potentiella patienter till kliniska studier.[31] Textanalyserande algoritmer har också visat sig kunna effektivisera arbetet med att hitta relevanta studier till vetenskapliga översiktsartiklar.[32]

Etiska och samhälleliga aspekter på AI i hälso- och sjukvård

Patientsäkerhet

Ökad patientsäkerhet är en av de främsta drivkrafterna bakom det stora intresset för AI i hälso- och sjukvården. Många studier har visat lovande resultat när det gäller att använda AI för att förbättra diagnostik, prognostik och medicinskt beslutsfattande. Samtidigt finns en ökad medvetenhet om att det också finns potentiella patientsäkerhetsrisker kopplade till AI.

Avvikande fall

En AI-algoritm lär sig inte fakta om världen utan fakta om den databas den tränats på (se ruta s. 5). När den ställs inför data och scenarier som avviker från träningsdata kan träffsäkerheten därför minska. I den kliniska miljön kan detta leda till lägre tillförlitlighet om en algoritm används på ”otypiska” patienter som inte varit representerade i träningsdata, på en patientgrupp som systemet inte varit avsett för eller om det sker förändringar i patientgruppen över tid. I något fall har det varit tillräckligt att bilder kommit från en annan skanner än träningsdata för att tillförlitligheten ska ha minskat.[33] Dynamiska algoritmer som justerar sina modeller i takt med att de stöter på nya fall skulle kunna motverka sådana effekter, men kan samtidigt leda till mindre kontroll över de data som algoritmen lär sig av, vilket kan

leda till andra oönskade effekter (se s. 6). Att vi i många fall inte vet hur en algoritm kommer fram till sina resultat (se ruta s. 9) kan göra det svårare att förutse sådana misstag och att förstå vad misstagen beror på när de väl inträffar.

Att skilja mellan samvariation och orsakssamband

AI-algoritmer är utformade för att upptäcka samband i träningsdata som kan användas för förutsägelser när de presenteras för nya data. I en klinisk miljö är det avgörande huruvida ett sådant samband återspeglar en orsaksrelation eller inte. Den distinktionen förmår dock dagens AI inte göra. En algoritm som skulle bistå läkare i att förutse risken att avlida av lunginflammation fann att patienter med astma hade lägre sannolikhet för detta än befolkningen i stort.[34] Detta kontraintuitiva fynd beror på att astmatiker med lunginflammation som regel läggs in på intensivvårdsavdelning direkt och därför har lägre mortalitet. Att algoritmer inte ”förstår” orsakssamband blir särskilt problematiskt med icke-transparenta modeller där det är oklart hur algoritmen kommer fram till sina resultat (se ruta s. 9).

Kompetensförlust

Redan i dag finns mycket av det kollektiva medicinska vetandet lagrat i medicinsk litteratur, register och olika hälsodatasystem, inte hos enskilda yrkesutövare. Denna utveckling kan accelereras om AI tar över fler och fler arbetsuppgifter i vården och vi allt mer förlitar oss på dess beslut. Risken finns att AI-system tar över rollen som lagringsplats för det kollektiva medicinska kunnandet och att det sker en kompetensförlust hos personalen.[35] Detta kan leda till patientsäkerhetsrisker om systemen skulle fallera. Icke-transparenta system som vårdpersonalen inte förstår kan förstärka denna risk.

Goda modeller kräver god datakvalitet

Avgörande för tillförlitligheten hos en AI-algoritm är kvaliteten på de data som den använder under inlärningen. Om träningsdata innehåller för få datapunkter eller om relevanta variabler saknas kan algoritmen inte förutsäga utfallsvärden med tillräcklig precision. Om exemplen i träningsdata i stället innehåller felaktig eller irrelevant information (”brus”) kan modellen bli alltför väl anpassad till träningsdata och leda till att systemet får en lägre träffsäkerhet när det ställs inför nya exempel. Ett exempel är en algoritm för att känna igen hudcancer, som visade sig använda de markeringar som läkare ofta gör kring misstänkta cancerfall som grund för diagnossättning.[36]

”Artificiell dumhet”

Under träningsfasen lär sig en algoritm att känna igen mönster i träningsdata. Vid bilddiagnostik får applikationen normalt gå igenom ett stort antal bilder för att hitta mönster som skiljer positiva från negativa svar. Men studier visar att det som en AI-algoritm ”känner igen” när den identifierar en viss typ av objekt kan vara något helt annat än det vi människor förknippar med samma objekt. Detta har visat sig kunna leda till uppseendeväckande fel, där sådant som vi skulle klassa som oidentifierbart brus uppfattas som objekt, eller där små, oväsentliga förändringar får objekt att omtolkas helt.[37] Fel av denna typ skulle kunna få allvarliga konsekvenser i hälso- och sjukvården.

”Automation bias”

Om AI-stöd används för allt fler av vårdens beslut finns risk för så kallad ”automation bias”, där personalen litar alltför mycket på automatiserade bedömningssystem och slutar leta efter evidens som bekräftar eller motsäger resultaten. Risken är särskilt stor om det är system som visat sig överlag tillförlitliga.[38]

Att balansera nytta och risk

Det sätt som algoritmer lär sig på gör att ett system kan ha hög tillförlitlighet i situationer som liknar dem det har tränats på, samtidigt som det kan göra uppseendeväckande fel om det stöter på data som avviker från träningsdata (se ruta ovan). Att använda AI i hälso- och sjukvården kan därför

leda till att många patienter får bättre och snabbare vård, samtidigt som några patienter utsätts för risker som inte skulle ha uppstått om läkare eller andra vårdanställda skulle ha gjort bedömningarna. En grundläggande fråga är hur risken för fel ska balanseras mot de vinster som AI i vården kan erbjuda. Är det acceptabelt att ett mindre antal patienter riskerar komma till skada om många patienter samtidigt får bättre vård?

Medikalisering och överdiagnostik

En ökad användning av AI-baserade hälsoappar och kroppsnära teknik kan leda till tidig upptäckt av sjukdom och ökad självständighet för patienten. Men ständig övervakning av en människas hälsostatus skulle också kunna leda till hälsostress och till medikalisering av problem som i grunden bottnar i psykosociala faktorer. Algoritmer som granskar journaldata hos ett större antal patienter i syfte att upptäcka sjukdom eller sjukdomsrisk kan precis som annan screening* resultera i överdiagnostik och överbehandling och skapa onödig oro hos friska personer.

Rättvisa och ickediskriminering

Precis som andra nya metoder och teknologier som antas kunna leda till förbättrad hälsa för befolkningen aktualiserar AI i hälso- och sjukvården frågor om hur en jämlik tillgång kan säkras, så att inte införandet av AI i vården riskerar leda till ökande hälsoklyftor, exempelvis mellan olika socioekonomiska grupper eller mellan olika delar av landet. Liksom när det gäller andra metoder i vården kan det finnas risk att utvecklare av kommersiella skäl koncentrerar sig på sjukdomar och tillstånd som många lider av, och att patienter med sällsynta sjukdomar inte lika lätt får tillgång till den nya tekniken.

Vid sidan av detta riskerar AI genom sin särskilda karaktär att medföra att nya former av ojämlikhet uppstår eller att befintliga hot mot jämlikheten förstärks.

Efterfrågestyrd i stället för behovsstyrd vård

En ökad användning av AI-baserade patientinriktade applikationer kan leda till tidig upptäckt av sjukdom och ökad självständighet för patienten. Men det kan också leda till en mer efterfrågestyrd vård när patienter allt oftare söker vård för symtom som en hälsoapp har upptäckt, kanske med en önskan om en specifik behandling som appen har föreslagit. Detta skulle kunna leda till att patienter med större behov trängs undan och till ojämlikhet mellan patienter som använder sådana applikationer och patienter som inte gör det. AI-baserade journalgranskningar för att upptäcka sjukdomsrisk kan gynna patienter med hög konsumtion av vård och därmed mycket data i journalen, något som ofta, men inte alltid, återspeglar behovet.

Datarelaterad bias

Många sjukdomar yttrar sig olika hos män och kvinnor eller yngre och äldre. Det finns genetiska variationer mellan olika etniska grupper som kan vara av betydelse för diagnostik och behandling. I kliniska studier är ofta

> "En grundläggande fråga är hur risken för fel ska balanseras mot de vinster som AI i vården kan erbjuda."

* Undersökningar där man letar efter tecken på sjukdom hos ett större antal symtomfria personer.

kvinnor och äldre personer underrepresenterade, medan genetiska databaser ofta har en övervikt av personer med europeiskt ursprung.[39] Om en algoritm i vården lär sig från en träningsdatabas där vissa patientgrupper är underrepresenterade kan det leda till att dessa grupper löper större risk för feldiagnostisering. Skälet är att det ger högre total träffsäkerhet om algoritmen anpassar sina förutsägelser till den största gruppen.[40] Ett exempel är hudcancer, som yttrar sig olika hos vita och färgade. En algoritm för att diagnosticera hudcancer som i huvudsak tränats på vita patienter riskerar därför att få sämre träffsäkerhet för färgade patienter.[41] En fråga blir om vi är beredda att acceptera sådana skillnader om det samtidigt medför bättre diagnostik och behandling för det stora flertalet patienter.

Även om variabler som ålder, kön, etniskt ursprung m.m. inte förekommer i träningsdata kan ojämlikhet eller diskriminering uppstå om variabler i indata samvarierar med sådana egenskaper. Inom hälso- och sjukvården skulle exempelvis användning av variabler som samvarierar med ålder kunna leda till att yngre patienter får en icke avsedd fördel i prioriteringssituationer. Detta eftersom lägre ålder i många fall ökar sannolikheten för ett positivt utfall.

Valet av utfallsvärde, det som algoritmen ska förutsäga, kan också innebära att ojämlikheter i samhället riskerar att reproduceras. Studier har visat att amerikanska läkare oftare misstar depression för schizofreni hos afroamerikanska än hos vita patienter.[42] En algoritm som lär sig diagnostisera sjukdom från tidigare diagnossättning riskerar att reproducera sådana misstag. Ett annat exempel är en amerikansk algoritm som används för att förutsäga framtida sjukdomsrisk och därmed behovet av förebyggande vårdinsatser. En studie visade att afroamerikanska patienter behövde vara sjukare än vita för att komma i fråga för samma förebyggande vårdinsats.[43] Skälet var att man valt prognostiserade vårdkostnader som mått på framtida sjukdomsrisk, något som missgynnar afroamerikaner som generellt får mindre kostsam vård än vita vid samma sjukdomsbörda.

Det faktum att många AI-system använder icke-transparenta algoritmer (se ruta s. 9) kan göra det svårare att upptäcka olika typer av datarelaterad bias och diskriminering och orsakerna bakom.

Undergrävd riskfördelningsprincip

AI-baserade analyser som utgår från patientdata skulle kunna ge allt mer precisa förutsägelser av framtida sjukdomsrisk. För patienten innebär det att förebyggande åtgärder kan vidtas. Men sådan information skulle också kunna vara intressant för försäkringsbolag för att kunna erbjuda lägre eller högre premie beroende på kundens framtida sjukdomsrisk. Detta kan undergräva den riskdelningsprincip som är grunden för alla försäkringar och leda till att vissa grupper i samhället får svårare att teckna sjukförsäkring.

Samhällets bedömning av huruvida en behandling är kostnadseffektiv och bör ingå i vårdutbudet bygger också på en form av riskdelning, eftersom såväl nyttan som kostnaden oftast varierar mellan olika patienter. Med hjälp av AI skulle hälso- och sjukvården kunna göra allt mer precisa förutsägelser om den enskilda patientens förutsättningar att tillgodogöra sig en viss behandling. Frågan skulle därmed kunna väckas om även patienter som har mindre chanser att bli hjälpta ska erbjudas behandlingen, eller om kostnaden för dessa patienter ska anses för hög i relation till den förväntade effekten.

Inbyggda värderingar

För att kunna bedöma olika behandlingsalternativ i vården krävs kunskap om de olika alternativens förväntade nytta och risk. Men avvägningen mellan förväntad nytta och risk är inte endast en kunskapsfråga utan också en fråga om *preferenser* eller *värderingar*. Är exempelvis en högeffektiv behandling med svåra biverkningar att föredra framför en något mindre effektiv med färre biverkningar? Uppvägs risken för komplikationer vid en viss operation av möjligheten att slippa ett livslångt vårdberoende?

När en algoritm som ger behandlingsrekommendationer designas kommer därför ofrånkomligen värderingar och preferenser av olika slag att byggas in, exempelvis genom valet av utfallsvärden som algoritmen ska optimera.* Frågan blir vems eller vilkas preferenser och värderingar som byggs in. Olika läkare kan göra olika bedömningar av vilken behandling som är den mest lämpliga i en given situation.[44] Patienter har ofta olika preferenser när det gäller exempelvis avvägningen mellan förväntad nytta och risk. Frågan blir hur sådana skillnader kan beaktas i en AI-stödd hälso- och sjukvård.

Det har också framförts att om AI-algoritmer ger konkreta rekommendationer och råd till vårdpersonal behöver de inkorporera de etiska principer (exempelvis att göra gott, inte skada, respektera självbestämmandet och vara rättvis[45]) som vi anser bör vägleda hälso- och sjukvården.[46] Men för att kunna tillämpas behöver etiska principer tolkas och vägas mot varandra, och frågan blir åter vems tolkningar och avvägningar som ska byggas in. Dessutom är det inte givet att etiskt övervägande alltid kan representeras av algoritmer, särskilt i situationer där det råder ett informationsunderskott och förväntade nyttor och risker inte är fullt kända.[47]

Missbruk och illasinnad användning

Oetisk design

En algoritm skulle medvetet kunna designas för att optimera något annat utfall än det som samhället anser önskvärt. En parallell kan dras till Volkswagenskandalen, där bilars datasystem programmerades för att ge bästa betyg vid miljötester snarare än minsta möjliga utsläpp. I stället för att optimera hälsoutfallet skulle en algoritm kunna designas för att leva upp till uppställda kvalitetskriterier utan att det innebär bättre vård, eller för att maximera vårdgivarens vinst, exempelvis genom att föreslå åtgärder som premieras i ersättningssystemet.[48] Bristen på transparens hos många algoritmer kan försvåra upptäckten av ”oetisk design”.

Manipulation

I takt med att allt fler funktioner i samhället utförs av AI-system kommer incitamenten för att försöka påverka systemen att öka. Eftersom funktionen hos en algoritm är beroende av de data den tränats på skulle ett sätt att påverka ett system kunna vara att via cyberattacker manipulera den databas som algoritmen lär sig från. Användare skulle också kunna – kanske via andra AI-algoritmer – försöka lära sig hur ett system fungerar för att kunna modifiera de värden som matas in i syfte att styra utfallet i önskad riktning. Inom hälso- och sjukvården skulle manipulation av AI-system kunna användas exempelvis för att kringgå behandlingsriktlinjer eller påverka ersättningen (gäller särskilt försäkringsfinansierade hälso- och sjukvårdssystem).[49]

Autonomi

Som konstaterats bygger en rekommendation från en AI-algoritm på medvetet eller omedvetet inbyggda värderingar. När alltfler både små och stora beslut helt eller delvis lämnas

* Även diagnostiska AI-system aktualiserar värderingsfrågor. Det kan exempelvis anses viktigare att algoritmen inte missar några sjukdomsfall än att den ibland felaktigt pekar ut en frisk patient som sjuk. För många ”falskt positiva” svar kan samtidigt ta resurser i anspråk som behövs för andra patienter, vilket också kan behöva vägas in.

över till algoritmer kan våra val styras på ett sätt som, ofta utan att vi märker det, riskerar undergräva vår makt över oss själva, våra liv och det samhälle vi lever i. Detta gäller även på vårdområdet. Om erbjudandet av vård styrs av algoritmer som räknar ut vilken behandlingsplan som antas bäst gagna patienten – och bäst motsvara patientens preferenser – kan det leda till en begränsning av patientautonomin. Det gäller särskilt om det är icke-transparenta system där vårdpersonalen inte kan förklara grunderna för valen (se ruta).

Att använda AI för att förutsäga framtida sjukdomsrisk genom breda analyser av patientdata väcker frågor om vilket samtycke som ska krävas av patienten och hur patientens bästa ur ett hälsoperspektiv ska vägas mot rätten att avstå från vård och säga nej till oönskad hälsoinformation.

Hälsoappar och kroppsnära teknik som bygger på AI kan leda till ökad kontroll över den egna hälsan och ökad autonomi. Men ständig övervakning av ens hälsostatus kan också leda till hälsostress, där oro för den egna hälsan tar över vardagen och i praktiken begränsar autonomin.[50]

Integritet

För att kunna utveckla AI-algoritmer i vården behövs patientdata i stora volymer, inklusive högkänslig information som anamneser, provsvar, patientrapporterade uppgifter, diagnoser m.m. Detta aktualiserar generella frågor kopplade till Big Data, som exempelvis hur företag ska kunna få tillgång till data från vårdgivare som de behöver för att kunna utveckla nya tillämpningar, samtidigt som patienternas integritet kan skyddas och rätten till kontroll över data som rör den egna personen säkras.

När det inte är nödvändigt att kunna spåra vem data kommer ifrån, exempelvis när en algoritm lärs upp, är grundtanken att algoritmer endast ska ha tillgång till anonymiserade data. Men med tillräckligt mycket anonymiserade data kan det vara möjligt att – med hjälp av AI – återidentifiera data.[51] Ett sätt att komma runt detta som föreslagits är så kallade syntetiska databaser. Dessa konstrueras utifrån verkliga personuppgifter på ett sätt så att väsentlig information på gruppnivå bevaras, men varje ”person” i databasen inte representerar någon verklig människa.[52]

Black box-problemet

Många företag betraktar AI-algoritmer som företagshemligheter som man inte delar med sig av.[53] Men genom sin komplexitet är algoritmer för djup maskininlärning och andra avancerade AI-algoritmer i sig själva icke-transparenta teknologier, där vi många gånger inte vet hur eller på vilka grunder en algoritm kommer fram till sina resultat. Bristen på transparens hos många AI-algoritmer brukar benämnas ”black box-problemet”.

Inom hälso- och sjukvården skapar black box-problemet utmaningar när det gäller bland annat patientsäkerhet, icke-diskriminering, autonomi, ansvar och förtroende. För att hantera sådana utmaningar efterlyser många ökad transparens hos AI-system.[54] Vissa experter har gått så långt som att förorda stopp för alla icke-transparenta algoritmer i offentlig verksamhet.[55] Andra menar att man bör satsa på att utveckla tekniker för att förklara hur icke-transparenta AI-system fattar beslut för att därigenom öka transparensen och minska risken för misstag.[56]

En invändning är att om vi ska kunna tillvarata AI:s fulla potential och inte bromsa utvecklingen – exempelvis när det gäller att diagnostisera sjukdomar med högre precision än vad människor förmår – måste vi göra avkall på transparensen och möjligheten att förstå grunderna för besluten. Skälet är att avancerade AI-algoritmer hanterar information på en helt annan komplexitetsnivå än vad mänskliga hjärnor förmår.[57] Somliga gör till och med gällande att riktigt avancerade algoritmer till sin natur är omöjliga att förklara.[58] Det avgörande är emellertid inte att vi förstår hur algoritmerna fungerar – det gör vi inte heller med alla läkemedel – utan att de är säkra och effektiva. Därför, sägs det, bör fokus inte ligga på transparens och förklaringar, utan på att validera algoritmernas tillförlitlighet i den kontext där de ska användas.[59]

Särskilda integritetsproblem aktualiseras inom neuroteknologi när AI kopplas ihop med mänskliga hjärnor via så kallade BCI (brain-computer interfaces). Sådan teknik har redan utvecklats för att styra proteser och tros kunna få stor betydelse för behandling av sjukdomar som epilepsi, schizofreni och förlamning. Men den skulle också kunna göra det möjligt att avkoda mentala processer och direkt manipulera hjärnan.[60]

Integritetsrelaterade frågor kan också uppstå om AI används för olika former av övervakning, exempelvis övervakning av patienter i hemmet för att förebygga sjukhusinläggningar.[61]

Ansvar

Vårdpersonal utför sitt arbete under yrkesansvar och kan hållas juridiskt ansvariga för felaktiga bedömningar och beslut. En ökad användning av AI i hälso- och sjukvården aktualiserar frågor om var det juridiska ansvaret bör ligga om en diagnos eller en behandlingsrekommendation från en algoritm visar sig vara felaktig. Är det rimligt att hålla vårdpersonalen ansvarig för bedömningar som gjorts av ogenomträngliga ”svarta lådor” vars processer ingen kan förklara (se ruta s. 9)? Om inte, vem ska i så fall hållas ansvarig? Programmeraren? Tillverkaren? Myndigheten som godkänt systemet? Eller sjukhuset som använt sig av det? Kan systemet självt tillskrivas ansvar? Kanske bör det juridiska ansvaret ses som delat mellan flera olika parter.

För att tydliggöra ansvarsfrågan anförs ibland att AI i hälso- och sjukvården bör fungera som ”beslutsstödjare” och inte ”beslutsfattare” (jämför diskussionen om augmented intelligence s. 2).[62] Om en algoritm ger en behandlingsrekommendation som inte följer gällande riktlinjer måste läkaren bedöma om den bör följas eller inte. En sådan hållning tydliggör ansvarsfrågan – beslutet är läkarens, som också är ansvarig för det.*

Det är emellertid inte säkert att den behandling riktlinjerna föreslår är den optimala eller att läkarens bedömning är bättre. Kanske har algoritmen upptäckt att det finns fall där utfallet blir bättre med en annan behandling än den rekommenderade.[63] Ett viktigt syfte med att införliva AI i vården är just att förbättra bedömningarna jämfört med rådande läge. Om det är en transparent algoritm och läkaren förstår hur den kommer fram till sina resultat kanske det går att värdera om en oväntad rekommendation bör följas eller inte. Men för att öka transparensen kan man behöva göra avkall på träffsäkerheten (se ruta s. 9). Om vi vill att vårdpersonal ska ta ansvar för beslut som bygger på AI-baserade rekommendationer måste vi kanske i vissa situationer välja algoritmer som är mindre träffsäkra.

Om AI-algoritmer blir allt bättre och de felaktiga rekommendationerna allt färre, kan man kanske komma till en punkt där läkare ska förväntas att följa en rekommendation. Det har anförts att i ett läge där en läkare förväntas följa en rekommendation från en algoritm måste den som designat algoritmen hållas ansvarig för eventuella felbeslut.[64] Ett argument för att hålla utvecklare ansvariga är att de har störst möjlighet att undvika skadliga utfall. En sådan princip kan samtidigt ha en dämpande effekt på utvecklingen av nya algoritmer.[65]

Förtroende

För att AI ska kunna implementeras framgångsrikt i hälso- och sjukvården och patienterna få del av de vinster tekniken kan ge, måste både patienter och vårdpersonal kunna lita på AI-systemen. Om algoritmernas bedömningar inte håller hög säkerhet, eller är mindre precisa för vissa patientgrupper, finns risk att förtro-

* Motsvarande situation kan uppstå för andra personalgrupper, exempelvis om en biomedicinsk analytiker får ett resultat från en AI-applikation som avviker starkt från det förväntade.

endet för AI sjunker hos patienter och hälso- och sjukvårdspersonal. Även om algoritmerna har hög generell tillförlitlighet kan förtroendet skadas om de ibland gör uppseendeväckande misstag som en professionell mänsklig bedömare inte skulle göra. Detta kan även skada förtroendet för hälso- och sjukvården generellt.

Patienter måste vidare kunna känna tillit till att deras data är skyddade. Ansvarsfrågorna måste vara utredda om vårdpersonalen ska våga följa rekommendationer från AI-algoritmer.

Frågan är också hur förtroendet påverkas om allt fler patienter får diagnos- och behandlingsförslag direkt från patientinriktade AI-applikationer, utan att någon människa är inblandad. Kommer läkares sakkunskap att värderas mindre högt, och kan det reducera förtroendet för vården? Förtroendefrågan aktualiseras också om de förslag patienter får från en app skiljer sig från dem de får från sin läkare, särskilt om läkaren inte förstår hur algoritmen fungerar och kan förklara hur den kom fram till sina förslag. Frågan blir vem patienterna kommer lita på mest.

Black box-problemet (se ruta s. 9) har även bäring på frågan om förtroende. Transparens och öppenhet med avseende på den data som används av en algoritm och den logik som styr den kan öka tilliten, bland annat för att det blir lättare att bedöma tillförlitligheten. Hemliga eller ogenomträngliga algoritmer kan omvänt minska tilliten och leda till att implementeringen av potentiellt kvalitetsförbättrande metoder bromsas. Alltför stor öppenhet verkar dock kunna ha motsatt effekt på förtroendet, och öppnar också för manipulation, vilket i sig kan sänka förtroendet.[66]

Det finns i dag många hälsoappar där patienter kommunicerar med vårdpersonal via video. Samtidigt blir det allt vanligare att patientinriktade AI-applikationer använder datorgenererade ”virtuella hälsoassistenter ” i gränssnittet mot patienterna, möjligen i förtroendeskapande syfte. När tekniken förfinas kan dessa bli allt svårare att skilja från verkliga människor. Detta väcker frågan om hur det påverkar tilliten till vården om det inte är klart om det man möter i en hälsoapp är en ”bot” eller en verklig vårdpersonal. Bör patienten ha rätt att alltid få veta om man interagerar med en människa eller en dator?

Relationer

En ökad användning av patientinriktade AI-applikationer kan påverka den traditionella relationen mellan patient och hälso- och sjukvårdspersonal. AI-baserade hälsoappar och kroppsnära teknik kan stärka patienters kontroll över och kunskap om sin sjukdom och skapa en mer jämbördig relation med personalen. Samtidigt kan vårdens personal allt oftare komma att möta patienter som kommer med färdiga diagnos- och behandlingsförslag som de fått från AI-applikationer. Detta kan leda till en ny rollfördelning mellan patient och vårdpersonal, där de senare i högre grad får rollen att kvalitetssäkra eller tolka den information patienten redan fått.

Genom att AI-system i vården tar över allt fler rutinmässiga uppgifter skulle mer av vårdpersonalens tid kunna frigöras till patientmöten. I ett längre perspektiv är det emellertid tänkbart att en större del av interaktionen inom hälso- och sjukvården sker med AI-system i stället för med människor. En fråga är vad det betyder för patienternas välbefinnande att inte få träffa en läkare eller en sjuksköterska och vad det betyder för deras förtroende för vården när de inte möter några människor som kan förmedla detta förtroende.

”Kommer läkares sakkunskap att värderas mindre högt, och kan det reducera förtroendet för vården?”

En annan följd, om automatiska system tar över allt mer av de uppgifter som hittills utförts av läkare och annan vårdpersonal, kan bli att den traditionella relationen mellan patient och vårdpersonal försvagas och den centrala relationen bli den mellan patienten och sjukvårdssystemet.[67] Detta kan i sin tur leda till att uppfattningen om professionellt ansvar i vården urholkas, vilket skulle kunna få negativa konsekvenser för patientsäkerheten.

Utmaningar för framtiden

AI i hälso- och sjukvården har stor potential att skapa nytta för både patienter och samhälle. AI kan bidra till säkrare diagnostik, tidigare upptäckt av sjukdom och mer personanpassad behandling. Patientinriktade applikationer kan ge patienter bättre kontroll över sin sjukdom och därmed ökad självständighet och livskvalitet. Ur ett samhälleligt perspektiv kan AI bidra till förbättrad folkhälsa och till ett mer effektivt resursutnyttjande. AI i hälso- och sjukvården innebär samtidigt flera betydande utmaningar, av såväl teknisk, juridisk som etisk art, vilka behöver uppmärksammas och hanteras för att den enskilde inte ska fara illa och för att inte förtroendet för teknologin ska skadas och leda till sämre förutsättningar för att ta tillvara möjligheterna med AI.

KVALITETSSÄKRING. Hur kan algoritmer avsedda att användas i hälso- och sjukvården kvalitetssäkras och hur kan det säkerställas att de databaser som algoritmer tränas och testas på är kliniskt relevanta och representativa för de patienter de ska användas för? Vilka processer ska finnas för godkännande och uppföljning? Hälso- och sjukvårdspersonalen ska enligt lag ge vård som står i överensstämmelse med vetenskap och beprövad erfarenhet, medan något motsvarande krav inte finns på teknisk utrustning. Detta regelverk kan behöva anpassas när det i allt högre grad är datasystem och inte människor som ger vården. Dynamiska system som kontinuerligt förändrar sina modeller skapar särskilda utmaningar när det gäller kvalitetssäkring.

TRANSPARENS. Icke-transparenta system kan undergräva förtroendet hos patienter och vårdpersonal och göra det svårare att upptäcka och åtgärda bias och misstag. Hur kan systemen göras mer transparenta? Hur bör avvägningen mellan transparens och tillförlitlighet göras? Ska olika grad transparens krävas beroende på om det exempelvis handlar om diagnostik, behandlingsrekommendationer eller vårdbeslut?

ANSVAR. Moraliskt ansvar förutsätter möjlighet och mandat att påverka och tillräcklig information för att värdera olika handlingsalternativ. Var ligger ansvaret om en AI-algoritm i hälso- och sjukvården gör en felaktig bedömning och detta leder till att en patient kommer till skada? Hur bör det juridiska regelverket vara utformat så att ansvaret kan utkrävas på rätt nivå? Bör det juridiska ansvaret kunna fördelas mellan flera olika aktörer?

RÄTTVISA OCH SOLIDARITET. Hur ser vi till att de hälsovinster som AI i hälso- och sjukvården kan medföra kommer hela befolkningen till del? Hur kan vi säkerställa att bedömningar som görs av AI-algoritmer i vården återspeglar rimliga värderingar och inte innebär särbehandling eller ojämlikhet? Hur kan den riskfördelningsprincip som den offentliga hälso- och sjukvården bygger på upprätthållas om man med hjälp av AI i kombination med stora mängder persondata kan göra allt mer precisa bedömningar av den enskildes risk för sjukdom?

Fortsättning ▼

AUTONOMI. Hur ska patienters värderingar kunna beaktas i en AI-stödd vård så att inte patientautonomin minskar? Vilka rättigheter ska patienter ha när det gäller att få algoritm-baserade bedömningar och beslut förklarade för sig, så att de kan fatta informerade beslut om sin vård? Ska patienten ha rätt till en andra medicinsk bedömning av en läkare? Ska man ha rätt att tacka nej till AI-vård?

INTEGRITET OCH DATASKYDD. Rätten till skydd för den personliga integriteten när allt fler uppgifter samlas in om enskilda medborgare är inte en fråga som endast berör AI. Givet den stora mängd känsliga patientdata som krävs för att utveckla AI i hälso- och sjukvården är det dock avgörande att frågor om dataskydd och rätten till kontroll över data som rör den egna personen får sin lösning, liksom hur data kan tillgängliggöras för utvecklare på ett etiskt hållbart sätt.

Slutord

AI i hälso- och sjukvården som utformas, introduceras och följs upp på ett genomtänkt sätt kan medföra nytta för både patienter och samhälle. Detta förutsätter medvetenhet om såväl de möjligheter som de risker som teknologin innebär. Det är viktigt att ta de etiska utmaningarna kopplade till AI på allvar och inte uppfatta dem som hinder för innovation, utan något som kan stimulera och vägleda utvecklingen i riktning mot tillämpningar som främjar gemensamma mål och värden.

Statens medicinsk-etiska råd
(S 1985:A)
smer@regeringskansliet.se
08-405 10 00
@smer_nyheter
www.smer.se

Smer vill rikta ett tack till Olle Häggström (Chalmers), Mikael Laaksoharju (Uppsala universitet) samt Stefan Larsson (Lunds universitet) för att de lämnat synpunkter på ett utkast till denna skrift.

Läs mer

AI4People. (2018). *AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.*

Academy of Royal Medical Colleges. (2019). *Artificial Intelligence in Healthcare.*

AMA Journal of Ethics. (2019;21:E119–197). Artificial Intelligence in Health Care.

EU-kommissionens expertgrupp på hög nivå för AI-frågor. (2019). *Etiska riktlinjer för tillförlitlig AI.*

European Group on Ethics in Science and New Technologies. (2018). *Artificial Intelligence, Robotics and 'Autonomous' Systems.*

Future Advocacy/Wellcome. (2018). *Ethical, Social, and Political Challenges of Artificial Intelligence in Health.*

Larsson S m.fl. (2019). *Hållbar AI. Inventering av kunskapsläget för etiska, sociala och rättsliga utmaningar med artificiell intelligens.*

Teknologirådet. (2018). *Kunstig intelligens – muligheter, utfordringer og en plan for Norge.*

Nuffield Council on Bioethics. (2018). *Artificial intelligence (AI) in healthcare and research* (Bioethics briefing note).

Socialstyrelsen. (2019). *Digitala vårdtjänster och artificiell intelligens i hälso- och sjukvården.*

Noter

1 World Economic Forum. (2017). *The Global Risks Report 2017 12th Edition*.

2 McCarthy J m.fl. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine,* 27:12–14.

3 Davenport TH och Glover WJ (2018, 29 juni). Artificial Intelligence and the Augmentation of Health Care Decision-Making. *Nejm Catalyst*.

4 Socialstyrelsen. (2019). *Digitala vårdtjänster och artificiell intelligens i hälso- och sjukvården*.

5 Ibid.

6 Davenport TH och Glover WJ (2018, 29 juni). Artificial Intelligence and the Augmentation of Health Care Decision-Making. *Nejm Catalyst*

7 Boggs W. (2017, 14 juni). Artificial intelligence may help doctors keep up with new research. *Reuters*.

8 Esteva A m.fl. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542:115–118; Liu Y m.fl. (2019). Artificial Intelligence–Based Breast Cancer Nodal Metastasis Detection: Insights Into the Black Box for Pathologists. *Arch Pathol Lab Med*,143:859–868; Khosravi P m.fl. (2019). Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ Digit Med,* 2:21; Hu L m.fl. (2019). An Observational Study of Deep Learning and Automated Evaluation of Cervical Images for Cancer Screening. *J Natl Cancer Inst*, 111:923–932; *Axis Imaging News* [webbplats]. (2019, 8 oktober). AI Detects Key Chest X-Ray Findings Within 10 Seconds; De Fauw J m.fl. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*, 24:1342–1350.

9 Liang H m.fl. (2019). Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med*, 25:433–438.

10 Hanai TA, Ghassemi MM och Glass JR. (2018). Detecting Depression with Audio/Text Sequence Modeling of Interviews. *Interspeech,* 2018–2522.

11 Obermeyer Z och Emanuel EJ. (2016). Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med,* 375:1216–1219.

12 Desautels T m.fl. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Med Inform.* 4:e28.; Weng SF m.fl. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One,* 12:e0174944.

13 Miotto R m.fl. (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Sci Rep*, 6:26094.

14 Rajkomar A m.fl. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*, 1:18.

15 Somashekhar SP m.fl. (2018). Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. *Ann Oncol*, 29:418–423; Komorowski M m.fl. 2018). The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*, 24:1716–1720.

16 Thompson RF m.fl. (2018). Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation? *Radiother Oncol*, 129:421–426.

17 Lamanna C och Byrne L. (2018). Should Artificial Intelligence Augment Medical Decision Making? The Case for an Autonomy Algorithm. *AMA J Ethics*, 20:E902–910.

18 Mesko B. (2017). The role of artificial intelligence in precision medicine. *Expert Review of Precision Medicine and Drug Development*, 5:239–241

19 *Vetenskap&Hälsa.* (2018, 13 november). Vem ska få nytt hjärta? Artificiell intelligens och läkare gör ofta olika val.; Yaekley RM och Saxena M. (2018, 3 september). How could artificial intelligence benefit emergency medicine? *AI Med* [webbplats].

20 Kassahun Y m.fl. (2016). Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *Int J Comput Assist Radiol Surg*, 11:553–568; Smith R. (2019, 7 februari). How Robots and AI are Creating the 21st-Century Surgeon. *Robotics Business Review*.; Gholami B, Haddad WM och Bailey JM. (2018, 24 september). AI Could Provide Moment-by-Moment Nursing for a Hospital's Sickest Patients. *IEEE spectrum*.; Kuziemsky C m.fl. (2019). Role of Artificial Intelligence within the Telehealth Domain. *Yearb Med Inform*, 28:35–40.

21 Powell A. (2017, 22 december). AI Is Fueling Smarter Prosthetics Than Ever Before. *Wired*.

22 Bouton CE m.fl. (2016). Restoring cortical control of functional movement in a human with quadriplegia. *Nature*, 533:247–250.

23 Newton J. (2017, 20 januari). Artificial Intelligence App Ada: Your Personal Health Companion. *Medical News Bulletin*.; *Babylon GP at hand.* (2019). See an NHS GP in minutes for free.

24 Sensely. [u.å]. *Introducing Ask NHS, powered by Sensely* [webbplats]

25 Twentyman J. (2017, 16 mars). Arthritis Research UK enlists AI chatbot 'Arthy' in mission to offer information and advice *Diginomica* [webbplats]; Labowitz D m.fl. (2017). Using Artificial Intelligence to Reduce the Risk of Nonadherence in Patients on Anticoagulation Therapy. *Stroke*, 48:1416–1419.

26 Javanmardian M och Lingampally A. (2018, 5 november). Can AI Address Health Care's Red-Tape Problem? *Harvard Business review*..

27 Chun A m.fl. (2000). *Nurse Rostering at the Hospital Authority of Hong Kong.* Joudaki H m.fl. (2015). Using Data Mining to Detect Health Care Fraud and Abuse: A Review of Literature. *Glob J Health Sci*, 7:194–202; Lichtl A. (2018, 11 januari). Telemedicine logistics: network optimization using artificial intelligence. *MedCityNews* [webbplats]; Zaidi D. (2018, 22 april). The 3 most valuable applications of AI in health care. *Venturebeat*.

28 Gershgorn D. (2017, 19 april). Artificial intelligence could build new drugs faster than any human team. *Quartz*.

29 Al-Lazikani B. (2013, 11 november). Artificial intelligence uses biggest disease database to fight cancer. *The Conversation*.

30 Williams K m.fl. (2015). Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. *J R Soc Interface*, 12:20141289.

31 Kent J. (2019, 18 juli). Artificial Intelligence Could Increase Clinical Trial Success Rates. *HealthITAnalytics*.

32 Wallace BC m.fl. (2017). Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach. *J Am Med Inform Assoc*, 24:1165–1168.

33 Challen R m.fl. (2018). Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 28:231–237

34 London AJ. (2019). Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Cent Rep*, 49:15–21; Caruana, R m.fl. (2015). Intelligible Models for HealthCare. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15,* doi:10.1145/2783258.2788613.

35 Char DS, Shah NH och Magnus D. (2018). Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *N Engl J Med,* 378: 981–983.

36 Winkler JK m.fl. (2019). Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition. *JAMA Dermatol*, 155:1135–1141.

37 Nguyen A, Yosinski J och Clune J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 427–436; Eykholt K m.fl. (2017). *Robust Physical-World Attacks on Deep Learning Visual Classification*.; University of Oxford, Department of computer science. [u.å.] *Are we safe in self-driving cars?*

38 Challen R m.fl. (2018). Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 28:231–237

39 Rothwell PM. (2006). Factors That Can Affect the External Validity of Randomised Controlled Trials. *PLoS Clin Trials*, 1:e9; Devlin H. (2018, 8 oktober). Genetics research 'biased towards studying white Europeans'. *The Guardian*.

40 Zou J och Schiebinger L. (2018). AI can be sexist and racist — it's time to make it fair. *Nature*, 559:324–326.

41 Adamson AS och Smith A, (2018). Machine Learning and Health Care Disparities in Dermatology. *JAMA Dermatol*, 154:1247–1248; Lashbrook A. (2018. 16 augusti). AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind. *The Atlantic*.

42 Gara MA m.fl. ((2018). A Naturalistic Study of Racial Disparities in Diagnoses at an Outpatient Behavioral Health Clinic. *Psychiatr Serv*. 70:130–134.

43 Obermeyer Z och Mullainathan S. (2019). Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70 Million People. I *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)* (s. 89-89). ACM.

44 Se exempelvis Ross C. (2017, 5 september). IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. *STAT*.

45 Beauchamp TL och and Childress JF. (2019). *Principles of Biomedical Ethics* (åttonde upplagan). Oxford University Press.

46 Char DS, Shah NH och Magnus D. (2018). Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *N Engl J Med,* 378: 981–983; Dignum V. (2017). Responsible Artificial Intelligence: Designing Ai for Human Values. *ICT Discoveries*, Special Issue No. 1.).

47 Sahlin N-E. (2018). It's values that matter. I Sahlin N-E (red.), *Science and Proven Experience: Johannes* (s. 77–86). Lund University Publications.

48 Char DS, Shah NH och Magnus D. (2018). Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *N Engl J Med,* 378: 981–983.

49 Finlayson SG m.fl. (2019). Adversarial attacks on medical machine learning. *Science, 363:1287-1289.).*

50 Smer (Statens-medicinsk-etiska råd). (2017). *Den kvantifierbara människan. Att själv mäta sin hälsa.* Smer rapport 2017:1.

51 La N m.fl. (2018). Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning. *JAMA Netw Open. 1:e186040.*

52 Bellovin SM, Dutta PK och Reitinger N. (2019). Privacy and Synthetic Datasets. *Stan Tech L. Rev*, 22.

53 Pasquale F. (2015). *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press.

54 Se exempelvis Förenta Nationerna. (2018). *World Economic and Social Survey 2018.*

55 Campolo A m.fl. (2017). *AI Now 2017 Report.*).

56 Turek M. (u.å.). Explainable Artificial Intelligence (XAI). *Defense Advanced Research Projects Agency* (Darpa); Local Interpretable Model-Agnostic Explanations (LIME): An Introduction. *O'Reilly*; Wachter S, Mittelstadt B och Russell C. (2018). Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard Journal of Law and Technology*, 31:841–887.

57 Weinberger D. (2018, 28 januari). Don't Make AI Artificially Stupid in the Name of Transparency. *Wired.*

58 Bornstein AM. 2016, 1 september). Is Artificial Intelligence Permanently Inscrutable? *Nautilus.*

59 London AJ. (2019). Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Cent Rep*, 49:15–21.

60 Yuste R m.fl. (2017). Four ethical priorities for neurotechnologies and AI. *Nature, 551:159–163.*

61 Smer (Statens medicinsk-etiska råd). (2014). *Robotar och övervakning i vården av äldre – etiska aspekter.* Smer rapport 2014:2.

62 Luxton DD. (2019). Should Watson Be Consulted for a Second Opinion? *AMA J Ethics, 21:E131–137.*

63 Se exempelvis Nemati S, Ghassemi MM och Clifford GD. (2016). Optimal medication dosing from suboptimal clinical examples: a deep reinforcement learning approach. *Conf Proc IEEE Eng Med Biol Soc,* 2016:2978–2981.

64 Martin K. (2018). Ethical Implications and Accountability of Algorithms. *J Bus Ethics*, doi.org/10.1007/s10551-018-3921-3.

65 Bartlett M. (2019, 5 april). Solving the AI Accountability Gap. *Towards Data Science.*

66 Hosanagar K och Jair V. (2018, 23 juli). We Need Transparency in Algorithms, But Too Much Can Backfire. *Harvard Business Review.*

67 Char DS, Shah NH och Magnus D. (2018). Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *N Engl J Med,* 378: 981–983.