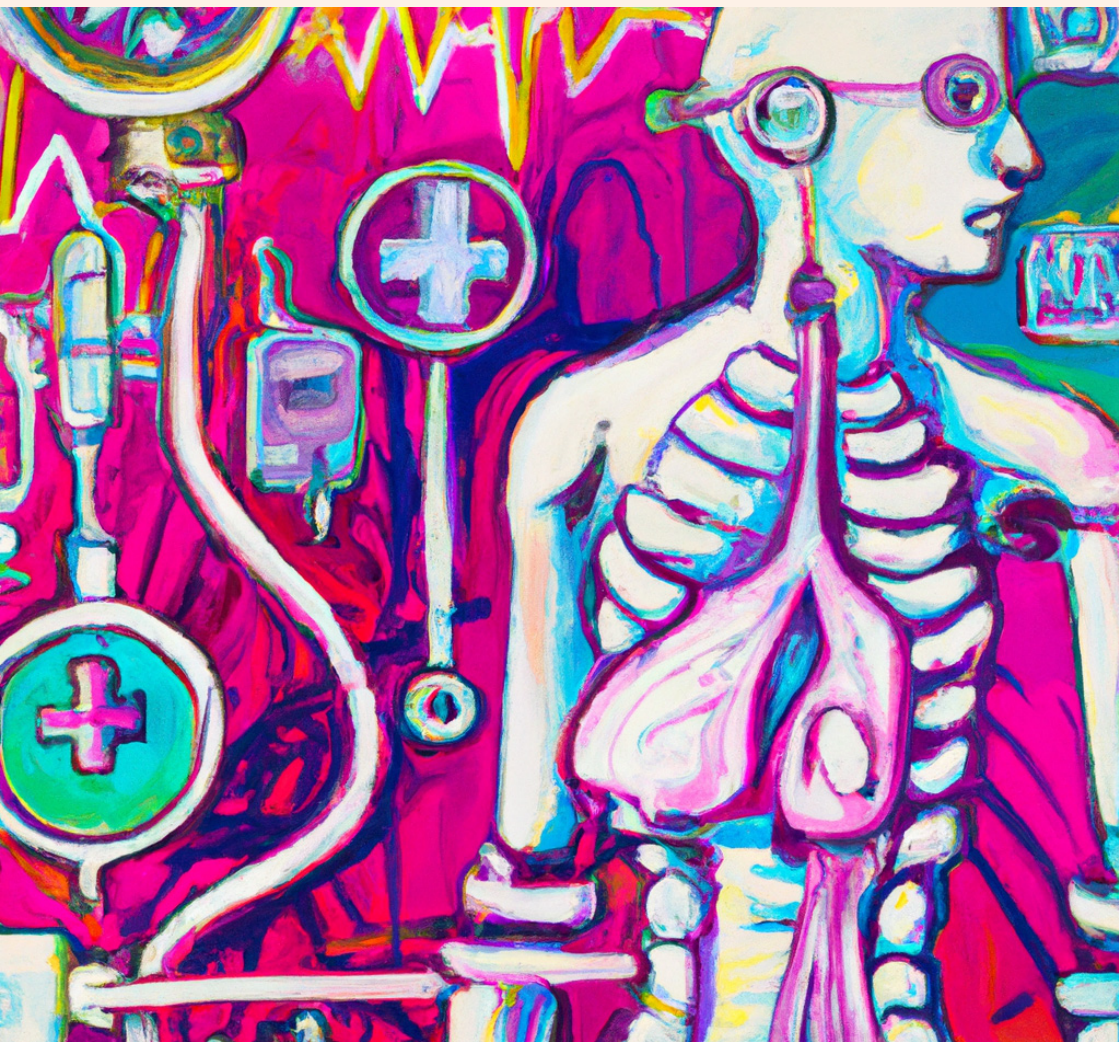




FORUM FÖR
HEALTH POLICY



Ett stort steg

Stora språkmodeller i sjukvården

Stockholm 2023

www.healthpolicy.se | info@healthpolicy.se

Forum för Health Policy är en ideell och politiskt obunden tankesmedja som vill stärka analysen av hälsopolitik i Sverige och stimulera till innovation och nya policyalternativ. Tillsammans med patientföreträdare, forskare och andra experter och med patientens behov i fokus och ett internationellt perspektiv är syftet att driva på innovation och utveckling i hälso- och sjukvård och social omsorg. Forum för Health Policy vill bidra med kunskap och möjliga policy- och handlingsalternativ för beslutsfattare.



Författaren: Jonathan Ilicki är legitimerad läkare, fellow i klinisk innovation vid KTH och Principal på Industrifonden. Författaren för denna rapport ansvarar själv för innehåll och slutsatser.

Illustrationer: Lynn Hsu (sida 29). Alla andra illustrationer har genererats av OpenAIs generativa bildtjänst DALL-E.

Tack till läkaren Amos Meir för allt stöd och intellektuell sparring under författandet av denna rapport.

Copyright © Jonathan Ilicki 2023

Innehåll

Förord	3
Sammanfattning	4
1. Sjukvård: tillämpa, kommunicera och samordna kunskap	8
2. Sjukvården har flera utmaningar	16
3. LLM kan adressera utmaningarna ..	30
4. LLM har begränsningar	44
5. Rekommendationer	56
6. Referenser	60

Förord

Forum för Health Policy har publicerat många rapporter med policyrekommendationer för att påskynda digital transformation som kan gynna sjukvården. Under det senaste året har det skrivits mycket om hur artificiell intelligens (AI) skulle kunna underlätta för sjukvården. Med tanke på den tilltagande bristen på sjukvårdspersonal har AIs potentiella roll blivit allt viktigare.

Forum för Health Policy har bjudit in läkaren Jonathan Ilicki att dela med sig av sina insikter och erfarenheter avseende AI i sjukvården. Denna rapport belyser hur AI skulle kunna öka patientsäkerheten och avlasta vårdpersonal, samt vilka risker AI innebär i vården, och syftar till att väcka diskussioner om AI:s roll i att transformera vården.

Vi vill tacka Jonathan Ilicki för hans värdefulla bidrag. Vi välkomnar dina tankar och feedback om detta ämne. Dela gärna dina kommentarer på vår hemsida eller via sociala medier.

Peter Graf

Ordförande, Forum för Health Policy
Stockholm, november 2023

Sammanfattning


Sjukvården har det tufft. Dokumentation och administration kräver en allt större del av vårdpersonalens tid. Kommunikation med patienter försvåras ofta av tidsbrist. Dessutom finns det stora utmaningar med att inhämta och tillämpa generell medicinsk kunskap så att den är relevant för patienter specifika situationer.

Stora språkmodeller (LLM) kan adressera dessa utmaningar och kommer därför att spela en viktig roll i hälso- och sjukvården. LLM kan frigöra tid för kliniker genom att automatisera dokumentation och viss kommunikation. De kan också sammanfatta stora mängder medicinsk litteratur för att ge vårdpersonal den medicinska kunskapen som är relevant för en specifik patient.

Som all ny teknik har LLM risker och begränsningar som måste hanteras. LLM kan hallucinera och ha fel. De kan också ha inbäddade och implicita värderingar vilket kan resultera i bias och diskriminerande output. Det är osannolikt att de kommer att kunna ersätta medmänskliga aspekter av sjukvård, såsom empati. Dessutom kan de även ge upphov till utmanande etiska avvägningar.

Trots dessa begränsningar testas LLM redan inom sjukvården. Sjukvården kommer att behöva ställa om för att dra nytta av LLM och samtidigt minimera riskerna. För det första behöver vissa vårdprocesser digitaliseras för att underlätta implementeringen av LLM. För det andra kommer de flesta vårdgivare att behöva utveckla nya kompetenser för att förstå och framgångsrikt implementera LLM. Slutligen är det viktigt att vårdgivare delar med sig av sina erfarenheter med LLM, då implementationerna ofta kommer att vara väldigt kontextberoende och lärdomarna kan vara högrelevanta för andra vårdgivare.

Förhoppningsvis kommer LLM att hjälpa vårdpersonal att lägga mindre tid bakom datorskärmarna och mer tid på att möta patienter där de är.



Jonathan Ilicki
November 2023



Sjukvård: att tillämpa, kommunicera och samordna kunskap

För att förstå hur stora språkmodeller kan påverka sjukvården måste vi först diskutera sjukvårdsprocesser - med fokus på centrala aktiviteter som återkommer över tid och i olika länder.

Att leverera sjukvård är komplext och kontextuellt och kan därför analyseras på olika sätt. Vissa sjukvårdsprocesser är dock universella och relevanta över tid. Denna rapport kommer att fokusera på tre kärnaktiviteter: att tillämpa, kommunicera och samordna medicinsk kunskap. Allt som påverkar dessa tre aktiviteter kommer alltså att kunna påverka sjukvården.

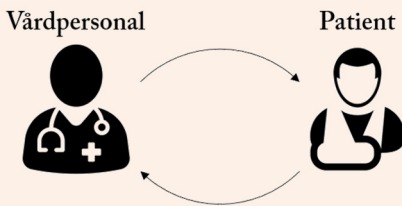
Dessa aktiviteter återfinns i alla sjukvårdssystem till följd av kunskaps- och färdighetsasymmetrin mellan patienter och vårdgivare. Vårdpersonal vet oftast mer än patienter om sjukdomar och behandlingar, och kan göra saker som patienter inte kan. Patienter skulle kunna ta lika väl hand om sig själva om det inte fanns någon asymmetri alls, och då skulle det inte finnas något större behov av att söka vård. Att förstå dessa aktiviteter i detalj är användbart för att förstå varför LLM kan ha en stor effekt på sjukvården.

1.1 Kommunicera kunskap

Kommunikation är centralt för all sjukvård. Detta omfattar vårdpersonal som ställer frågor till patienter, lyssnar på svar, svarar på frågor samt förklarar sina bedömningar eller annan information. 40-50% av klinikers tid ägnas åt att kommunicera med patienter, dock med stor variation mellan olika specialiteter.^{1,2}

Ju mer medicinsk kunskap sjukvården har, desto mer standardiseras kliniska processer med avseende på vilka frågor som ska ställas och hur patienter ska bedömas. Under de senaste decennierna har antalet och omfattningen av kliniska riktlinjer ökat påtagligt.³ En allt större del av samtalen bör vara och är standardiserade och innehåller därför standardiserade frågor eller diskussioner.

Figur 1.1. Kärnaktivitet: kommunikation



1.2 Tillämpa generell kunskap

Den andra återkommande processen är att vårdpersonal lär sig av, syntetiserar och tillämpar generell medicinsk kunskap. Medicinsk forskning och vetenskap genererar allmän kunskap, som vårdpersonal sedan behöver anpassa och omtolka till den specifika patienten som de ska handlägga. Över tid ansamlas alltmer kunskap, med en ökad förståelse för människokroppen och dess åkommor.

Publicerad medicinsk kunskap är som regel *generell* då den har flera begränsningar i hur den kan tillämpas. Exempelvis kan en studie på ett visst läkemedel visa att den har genomsnittlig effekt X på de patienter som var med i studien. Men denna kunskap garanterar *inte* att vi kommer att se exakt effekt X i en annan patientgrupp, som kan ha andra grundsjukdomar eller reagera annorlunda på läkemedlet. En viktig process inom sjukvården är därför att utgå ifrån allmän kunskap, och sedan omtolka och anpassa den innan den tillämpas för ett specifikt patientfall.

Första steget i denna omtolkning underlättas ofta av internationella, nationella eller lokala riktlinjer, som gör det lättare för kliniker att veta hur man bör hantera ett visst tillstånd. Den allmänna kunskapen behöver dock vägledas av klinisk expertis och förståelse för patientens unika situation och är en grundläggande komponent i evidensbaserad medicin. David Sackett, en av pionjärerna inom evidensbaserad medicin, beskriver det så här:

“Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough. Without clinical expertise, practice risks becoming tyrannised by evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient. Without

current best evidence, practice risks becoming rapidly out of date, to the detriment of patients.”⁴

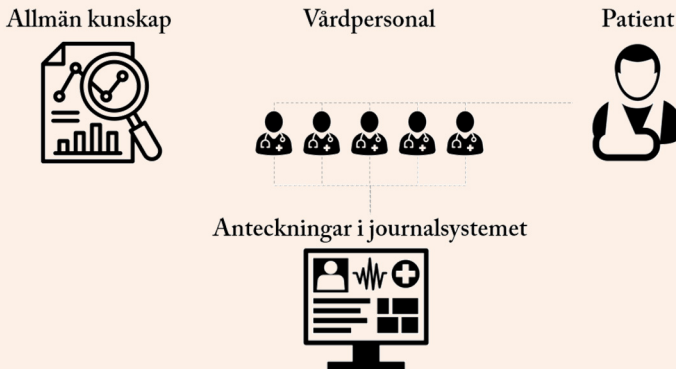
Figur 1.2. Kärnaktivitet: tillämpa och anpassa allmän kunskap efter ett specifikt patientfall



1.3 Samordna vård

Den tredje ständigt närvarande aktiviteten är att samordna vården av en patient, vilket görs till stor del genom journalanteckningar, som oftast lagras i ett elektroniskt journalsystem. Att samordna en patients vård kräver kännedom om tidigare kontakter med vårdgivare, andra behandlingsplaner eller operationer och att säkerställa en gemensam kunskapskälla för alla inblandade kliniker. Behovet av samordning och dokumentation ökar i takt med att patienter har alltmer interaktioner med sjukvården under en livstid, och att alltmer patientdata ackumuleras över tid. Antalet vårdpersonal har ökat per capita under de senaste decennierna, vilket sannolikt bidragit till ett ökat behov av samordning.⁵

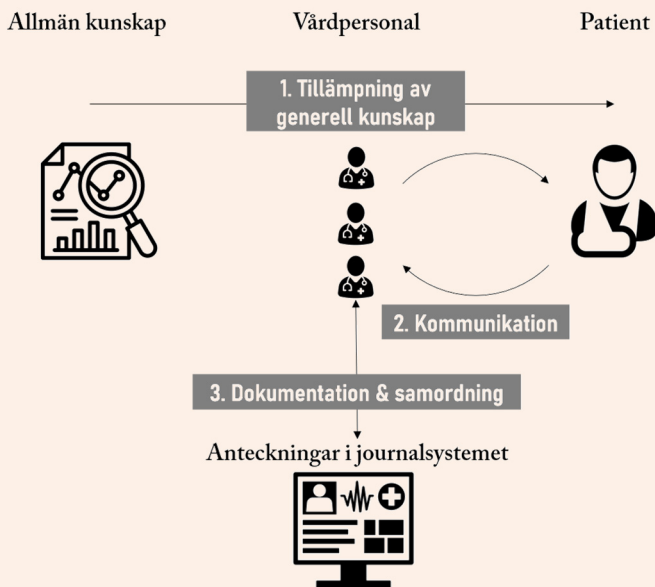
Figur 1.3. Kärnaktivitet: samordna vårdgivande genom dokumentation



1.4 Kärnaktiviteterna är sammanlänkade

Även om kärnaktiviteterna diskuteras separat så är de tätt sammanlänkade med varandra. Patientkommunikation är grundläggande för att veta vilken allmän kunskap man ska utgå ifrån, och allmän kunskap vägleder vårdpersonal om vad man ska fråga om. Att inhämta information från journalsystemet påverkar på vilken allmän kunskap man ska utgå ifrån och vad man ska diskutera med patienten. På samma sätt påverkar allmän kunskap och patientkommunikation vad som dokumenteras i en journalanteckning.

Figur 1.4. Tre universella och centrala sjukvårdsaktiviteter





Sjukvården har flera stora utmaningar

Sjukvården står inför flera stora utmaningar. I detta kapitel kommer vi att fokusera på några stora utmaningar som förekommer i alla utvecklade sjukvårdssystem, samt utmaningarnas underliggande orsaker.

2.1 Vårdpersonal saknas

Globalt saknas det cirka 10-15 miljoner läkare enligt WHO.⁶ Denna siffra kan verka abstrakt, men upplevs i många länder i form av långa köer till sjukvården, medicinska avvikelser och utbränd sjukvårdspersonal. Varför finns denna brist? Ett sätt att förstå bristen är att titta på de underliggande faktorerna som driver sjukvårdsanvändningen.

Inget enskilt nyckeltal kan fånga den exakta efterfrågan på sjukvård i ett land, men några nyckeltal kan indikera den totala efterfrågan. Några av de viktiga nyckeltalen är antalet invånare i ett land, medellivslängden (då man konsumerar mer sjukvård när man blir äldre) samt det tillgängliga medicinska utbudet av utredningar och behandlingar. På samma sätt kan man inte exakt mäta ett lands kapacitet att leverera sjukvård, men ett lands totala sjukvårdsutgifter säger något om hur mycket sjukvård som tillhandahålls.

Om man sammanfattar de förenklade idéerna ovan kan man illustrativt beskriva det som att:

- Totala efterfrågan på sjukvård = Befolkningsmängd × Genomsnittlig livslängd × Medicinskt utbud
- Totala tillgången på sjukvård = Hälso- och sjukvårdsutgifter.

Historiskt har tillgången ökat, vilket avspeglats i växande sjukvårdsutgifter.⁷ Efterfrågan för sjukvård har dock ökat ännu fortare, till följd av en åldrande befolkning⁸, ökningarna i genomsnittliga livslängden⁹ samt en mycket bredare utbud av medicinska behandlingar och utredningar (både i termer av vad som går att göra, och vad medicinska professionen och samhället anser bör göras).¹⁰ Dessa nyckeltal beskrivs för Sverige i tabellen nedan.

Tabell 2.1. Sveriges ökande sjukvårdsefterfråga & -tillgång 1971-2021.

	1971	2021	Förändring
Befolkning ¹¹	8.1M	10.4M	+28%
Genomsnittlig livslängd ¹²	74.5*	82.5*	+11%
Utbud av behandlingar och utredningar	Utgångsläge	Påtagligt utökat	Kan ej beskrivas kvantitativt [†]
Sjukvårdsutgifter per capita (\$/år) ¹³	306	811 [‡]	+165%

*77 för kvinnor, 72 för män; ökade till 84 och 81 i 2021. [†]Not:

nyckeltal för efterfråga ska multipliceras med varandra. Om utbudet av medicinska behandlingar hade ökat med t.ex. 20%, så skulle den totala ökningen vara $1.28 \times 1.11 \times 1.20 = 1.71$, det vill säga mer än ökningen i tillgång. Detta är en grov förenkling, men illustrerar att faktorerna inte bara är additiva. [‡]Inflationsjusterad siffra (nominell är \$6 228)

Faktorerna som diskuteras ovan är på intet sätt uttömmande eller alltomfattande. Andra faktorer, såsom Baumoleffekten och minskande risktolerans över tid, har också stor bäring på denna jämvik, men tas inte upp här.

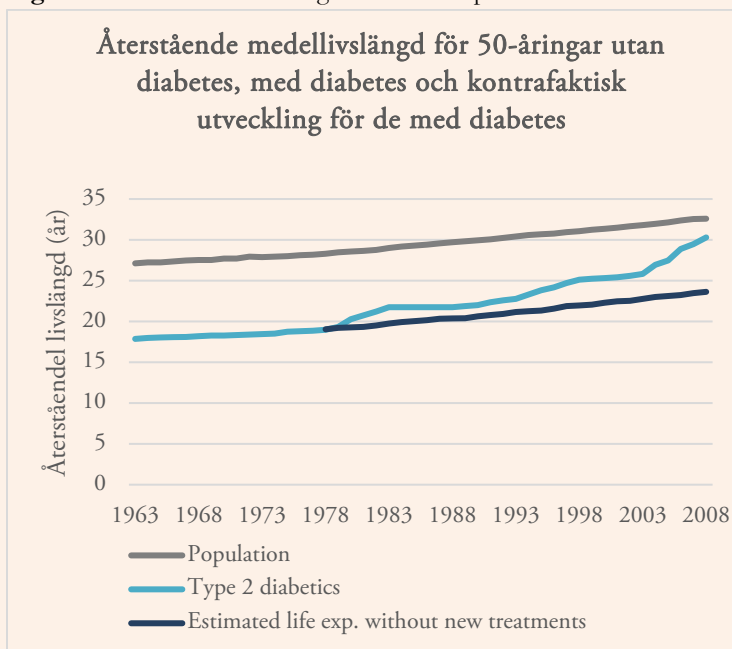
2.2 Rotorsakerna till ökad efterfråga är positiva

En ökad efterfråga på sjukvård utmanar sjukvårdssystemet, men det är viktigt att komma ihåg att de bakomliggande orsakerna är positiva och lovvärda.

1. Längre medellivslängd: Bättre behandlingar ökar medellivslängden. Moderna behandlingar förlänger livet för många patienter, t.ex. för cancer¹⁴, diabetes¹⁵ och infektionssjukdomar¹⁶. Många av förbättringarna i medellivslängden beror dock sannolikt inte på sjukvården, utan snarare på satsningar på folkhälsa, förbättrad hygien och infrastruktur.¹⁷ Det verkar dock troligt att ungefär hälften av ökningen i medellivslängd i modern tid har uppstått till följd av förbättringar i sjukvården.^{18, 19}

Ju bättre vården blir, desto mer arbete finns det att göra: När sjukvården är framgångsrik lever patienter allt längre, och det betyder att det blir allt fler patienter att följa upp över en längre tid.

Figur 2.2.1. Förväntad livslängd över tid för personer med diabetes¹⁵

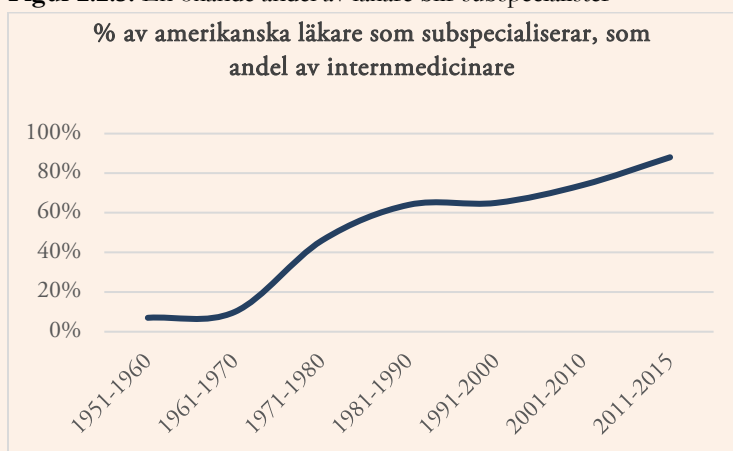


2. Utökad medicinsk kunskap: Mer medicinsk kunskap leder till nya behandlingar och större medicinsk omfattning (att man kan och vill behandla sådant som tidigare inte behandlats). Detta ger upphov till fler riktlinjer att följa.^{3, 10}

Ju mer medicinsk kunskap, desto mer finns det att ta hänsyn till: Vi vet i allt högre grad hur vi ska behandla patienter på bästa sätt, men denna ökning i information kan överbelasta vårdpersonal med instruktioner och riktlinjer.^{20, 21}

3. Vi har fler experter: Ökad medicinsk kunskap medför ett allt större värde av specialisering och subspecialisering. Specialiseringen började accelerera under slutet av 1900-talet i flera sjukvårdssystem. I USA ökade antalet subspecialister inom internmedicin från 7% på 1950-talet till 88% på 2010-talet.²²

Figur 2.2.3. En ökande andel av läkare blir subspecialister²²



Subspecialisering möjliggör mer specialiserad vård, men kräver mer samordning, ofta via journalanteckningar i journalsystemet. Införandet av elektroniska journalsystem har bidragit till bättre vårdkvalitet (genom att förbättra samordningen och minska vissa medicinska risker) men ofta till priset av ökad administration och lägre produktivitet.²³⁻²⁵

Ju mer specialisering, desto större behov av samordning och dokumentation: När vården är framgångsrik kan vi ge en allt effektivare vård, men det ökar samtidigt behovet av att samordna, dokumentera och läsa medicinsk information i journalsystemet.

Dessa tre grundorsaker illustrerar ett viktigt men ofta förbiset förhållande: ju bättre sjukvården är på icke-kurativa och icke-förebyggande behandlingar, desto större blir efterfrågan på sjukvård.

2.3 Kärnaktiviteterna har utmaningar

Utmaningarna med en växande och åldrande befolkning, en ökad medicinsk kunskap och ökad specialisering slår direkt mot sjukvårdens kärnaktiviteter: att kommunicera, tillämpa och samordna kunskap.

1. Kliniker saknar tid för att kommunicera med patienter: Att kommunicera med patienter är svårt om man inte har tid. När patienterna ökar i antal och komplexitet behöver mer information utbytas, vilket tar mer tid. Vissa analyser indikerar att det skulle ta cirka 7 timmar per dag för en läkare att följa alla förebyggande riktlinjer som fanns år 2003. Nyare analyser är närmare 14

timmar per dag.^{21, 26} Idag skulle en allmänläkare i USA behöva 27 timmar per dygn för att implementera och dokumentera alla relevanta riktlinjer för de patienter de möter.²⁰ Allmänläkare saknar tid att fråga alla som borde frågas och säga allt som borde sägas.

Figur 2.3.1. Omöjligt att följa alla riktlinjer^{21, 26, 27}



I en studie besvarades medicinska frågor från ett offentligt forum av läkare samt av en LLM-chatbot.²⁸ Dessa svar jämfördes sedan med varandra, och bedömdes avseende kvalitén på informationen och hur empatiskt svaret var. I 79% av fallen föredrog utvärderarna chatbotssvaren, som överträffade läkarnas svar. Författarna påpekar dock att detta delvis kan bero på längden på svaren. Chatbotens svar hade i genomsnitt fyra gånger så många ord. Längre läkarsvar föredrogs oftare och fick högre poäng av utvärderarna. Men detta återspeglar delvis verkligheten: till följd av begränsad tid måste läkare skriva korta och koncisa svar.

Figur 2.3.2. Mänskliga utvärderare bedömde att en chatbot gav mer empatiska och högkvalitativa svar än mänskliga läkare²⁸



2. Omöjligt att komma ihåg och tillämpa all relevant kunskap: Vi har enorma mängder kunskap om hur man bäst diagnostiserar och behandlar patienter. Exempelvis har *Canadian Medical Association* en databas med över 1700 riktlinjer för kliniskt bruk.²⁹ Men på grund av den oerhörda volymen är det oerhört svårt för vårdpersonal att ha alla riktlinjer i åtanke, tillämpa allt som är relevant och behandla alla patienter på ett optimalt sätt.²⁰

Medicinska misstag är en av de vanligaste dödsorsakerna i sjukvården och uppskattas orsaka mellan 250 - 800 000 dödsfall i USA årligen.^{30, 31} 10-15% av kliniska beslut uppskattas vara felaktiga (men alla uppskattningar av den karaktären är spekulativa då det kan vara svårt att definiera vilka beslut som är rätt eller fel).³² Däremot vet vi att läkare ofta beställer onödiga prover före operationer³³ eller i samband med canceruppföljning,³⁴ och har svårt att tillämpa den bästa kunskapen när man bedömer sannolikheten att en patient har ett visst tillstånd.^{35, 36} Inte bara det; kliniker misslyckas ofta med att uppskatta om en patient kommer att ha nytta av en viss behandling.^{35, 37} Att uppskatta och värdera sannolikheter för olika tillstånd och behandlingseffekter är ofta kontraintuitivt, och den mänskliga hjärnan är inte utformad för att memorera stora mängder data och sannolikhetskvoter. Denna informationsöverblastning, i kombination med snabbt

medicinsk ökande komplexitet, kan bidra till medicinska misstag.

En majoritet av frågor som uppstår i den kliniska vardagen verkar kunna besvaras med allmän och publicerad medicinsk kunskap, men vissa (och som regel viktiga) frågor kräver en syntes av allmän och patientspecifik data.^{38, 39} Den allmänna litteraturen kan alltså ofta besvara frågor som dyker upp i det dagliga kliniska arbetet, men att hitta och få tillgång till denna kunskap är ofta för tidskrävande för att det ska vara genomförbart.⁴⁰

3. Dokumentation och administration tar mycket tid:

Dessvärre tar det idag mycket tid att dokumentera patientarbete, och det genererar även stora mängder överflödigt dokumentation som tar ännu mer tid att läsa och navigera. Många kliniker frustreras över att dokumentation tar så mycket tid.⁴¹ Tiden varierar mellan specialiteter och länder, men det verkar som om ca 30% av många klinikers tid går åt till att dokumentation.^{2, 42, 43}

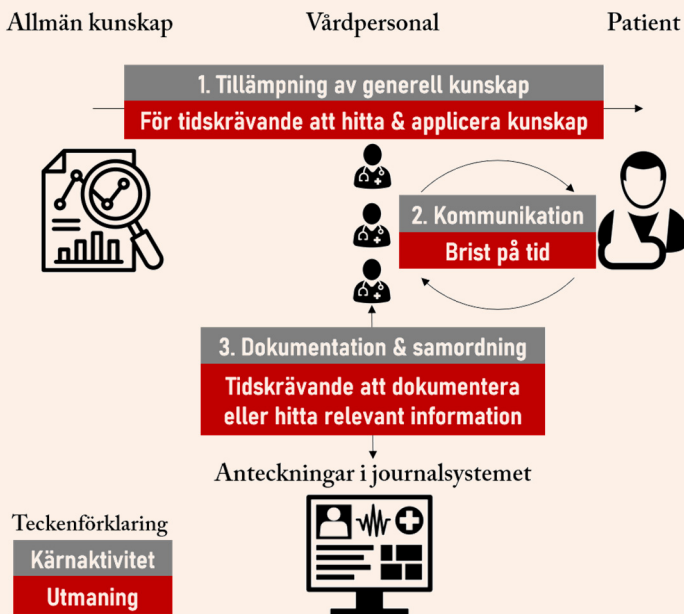
Journalinnehåll upprepas även ofta, vilket bidrar till en onödigt stor mängd text som tar mycket tid att läsa.⁴⁴

Detta skiljer sig mellan olika sjukvårdssystem och beror på hur man räknar, men 30-50% av journalanteckningar verkar ofta vara upprepningar.^{45, 46} I en svensk studie som

tittade på 98 journaler hade alla journaler upprepningar, och en journal hade en anteckning som hade kopierats 16 gånger.⁴⁷ En annan analys av 30 patientjournaler hittade 822 fall av upprepningar, med upprepningar i alla journaler.⁴⁸ Information behövs för att tillhandahålla god vård, men att konstant söka igenom text med upprepningar tar tid från patientmötet och gör det svårt att hitta relevant information.

Sjukvården står alltså inför stora utmaningar inom alla tre kärnaktiviteterna. Dessutom fortsätter de underliggande rotorsakerna till utmaningarna, så det förefaller osannolikt att dessa utmaningar skulle avstanna av sig själva.

Figur 2.3.3. Utmaningar för sjukvårdens kärnaktiviteter





LLM kan adressera flera av utmaningarna i sjukvården

En stor språkmodell (LLM) är en algoritm som har tränats på en stor mängd text. Detta gör att den kan tolka såväl som generera text med hög precision och komplexitet. Denna typ av AI kan adressera flera av utmaningarna som sjukvården står inför.



“Please fill out these medical forms, which are identical to the ones you filled out earlier online, and have the exact same questions your doctor will ask you later in the exam room.”

3.1 LLM kan automatisera repetitiv och standardiserad kommunikation

Kommunikationen med patienter blir alltmer standardiserad i takt med att den medicinska kunskapen ökar. Det betyder inte att vårdpersonal borde kommunicera med alla patienter på samma sätt, utan att vissa frågor alltid ska ställas när man utreder ett visst

tillstånd, eller att viss information alltid ska ges före en given behandling. En allt större del av kommunikationen blir alltså standardiserad desto vi bättre vi förstår hur olika tillstånd ska handläggas.

LLM kan hjälpa vårdpersonal att återfå tid genom att hjälpa till med repetitiv, generisk och standardiserad kommunikation. LLM i form av chatbots kan samla in standardiserad information från patienter genom ett automatiserat samtal. Idag kan digitala triage-botar redan spara en betydande mängd tid för vårdpersonal^{49, 50} I Sverige har en digital chatbot kunnat minska tiden för administrativa ärenden med så mycket som 68%.⁴⁹ LLM kan komplettera sådana chatbotar. Genom att både samla in och förmedla standardiserad information inför möten med vårdpersonal, kan de frigöra tid för ett mer patient-centrerat samtal.

3.2 LLM kan underlätta tillgången till relevant kunskap

Det är något vackert med hur mänskligheten år efter år samlat på sig medicinsk kunskap. Mänskligheten har genom århundraden arbetat hårt – inte bara med att förstå hur kroppen fungerar och hur man bäst kan hjälpa

patienter – utan även att dokumentera och tillgängliggöra denna kunskap i publikationer och läroböcker. Det dröjer dock som regel lång tid innan ny kunskap når kliniskt bruk, och många refererar till ett gap på 17 år från när ny kunskap finns tills att den når patienter.⁵¹ Under dessa år levereras suboptimal vård – vilket gynnar ingen.

LLM låter oss brygga gapet mellan vad vi gör och vad vi borde göra. LLM kan idag svara på medicinska frågor med hög noggrannhet, och har redan uppnått 85% korrekta svar på MedQA-datasetet (vilket innehåller medicinska tentafrågor, frågor från patienter samt frågor om medicinsk forskning).^{52, 53} På ett annat dataset (PubMedQA) uppnår LLM över 80% korrekta svar (människor uppnår cirka 78%). LLM som är fritt tillgängliga kan svara så pass korrekt att de skulle bli godkända på den slutgiltiga läkartentan i Polen.⁵⁴

Denna utveckling har varit exceptionellt snabb, men det är viktigt att komma ihåg att sjukvård handlar om många andra saker än att enbart svara korrekt på medicinska frågor. Många aspekter av ett kliniskt sammanhang är svåra för LLM att hantera, till exempel att väga in patienters implicita preferenser eller kulturella kontext.

LLM kan dock erbjuda alla kliniker en virtuell kollega att rådfråga – som kan svara pedagogiskt, ta hänsyn till den senaste forskningen och patientens egenskaper, och hjälpa vårdpersonal med kliniska frågor. Detta har aldrig tidigare varit möjligt: att bokstavligen talat dra nytta av all publicerad medicinsk kunskap som mänskligheten ansamlat över åren - och göra den enkelt tillgänglig för alla. Detta innebär ett unikt sätt att förbättra vårdkvalitet, och hjälpa kliniker att undvika både överdiagnostik och överbehandling, samt undvika att missa tillstånd som inte bör missas. Inte genom att arbeta hårdare, utan genom att använda den kunskap som tidigare generationer av kliniker och forskare har gett oss.

AI kan rädda liv genom att använda befintlig kunskap för att föreslå nya behandlingar

Every Cure är en ideell organisation vars mål är att rädda liv genom att använda befintliga läkemedel för att behandla nya tillstånd. De har utvecklat Linkmap, en AI-algoritm som poängsätter alla befintliga FDA-godkända läkemedels potential att behandla 12 000 mänskliga sjukdomar (totalt 36 miljoner utvärderingar). Denna AI-applikation har redan räddat liv. En patient som led av idiopatisk multicentrisk Castlemans sjukdom (iMCD), en sällsynt och livshotande sjukdom, hade inga återstående behandlingsalternativ och var på väg att läggas in på hospice. AI-algoritmen identifierade en möjlig alternativ behandling med ett redan existerande läkemedel som godkänts för andra tillstånd. Efter påbörjad behandling förbättrades patienten och symtomen avtog.⁵⁵

3.3 LLM kan automatisera dokumentation och tillgängliggöra relevant patientdata

LLM kan avsevärt minska tiden för dokumentation, samt automatiskt sälla fram anteckningar med relevant medicinsk information. Idag finns det redan tekniska lösningar som kan spela in ett samtal mellan patienter och vårdpersonal, och därefter transkribera och skriva om detta till en journalanteckning med hög noggrannhet.^{56, 57} Vissa vårdgivare och leverantörer har beskrivit att sådana system kan minska 75% av tiden som läggs på dokumentation.^{58, 59} LLM kan också ta till sig journalanteckningar och svara på frågor om dess innehåll.^{57, 60} Detta kan göra det mycket lättare för vårdpersonal att hitta relevant information i långa journaler. På detta sätt kan LLM minska tiden som läggs på att dokumentera och läsa i journalen, samtidigt som vi kan fortsätta koordinera vård i ett allt mer komplext sjukvårdssystem.

3.4 LLM har redan en otrolig prestanda och kommer bara att bli bättre

Snabb utveckling

Stora språkmodellers prestanda har ökat raskt under de senaste åren. Nyligen genomfördes det hittills största Turing-testet där människor skulle gissa på om de chattade

med en chatbot eller en människa. Människorna som pratade med chatboten gissade bara rätt på huruvida de pratade med en människa eller bot 60% av gångerna, vilket inte är mycket bättre än slumpen.⁶¹

Figur 3.4.1. Exempel på chattar från “*Human or Not?*” Turing-testet⁶¹.

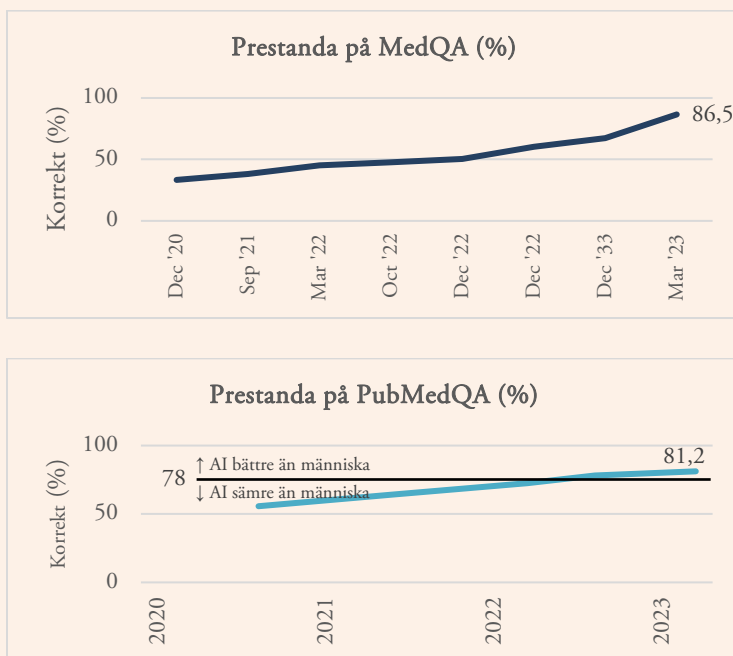


Bilden ovan visar två chattsamtal från Turing-testet.

Försök att gissa vilka av deltagarna som är en AI-chatbot eller människa, och reflektera över hur säker du är på ditt svar. Facit hittar du i referenskapitlet.

LLM:s medicinska prestanda har förbättrats snabbt. På några år har deras förmåga att besvara medicinska frågor förbättrats från att vara ytterst begränsad till att vara högpresterande tvärs flera olika typer av tester och frågor. Denna förbättring har inte bara skett för allmänna medicinska frågor, utan även för detaljerade frågor inom medicinska specialiteter.⁶²⁻⁶⁴

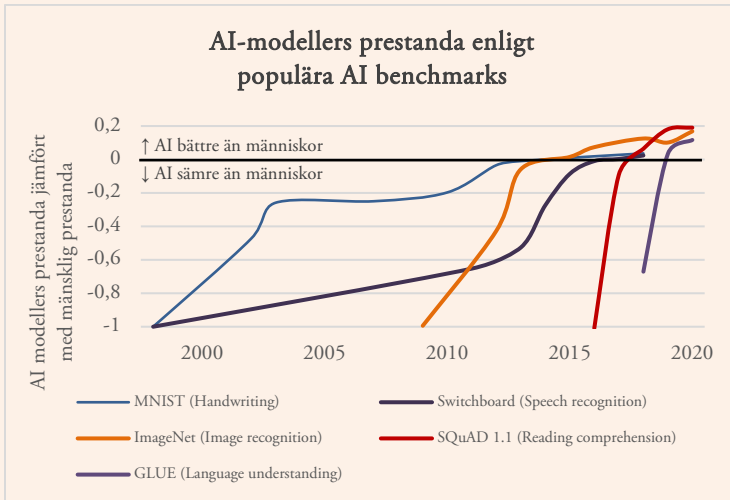
Figur 3.4.2. LLM-prestanda har förbättrats på flera medicinska dataset under de senaste åren^{53, 65}



Desutom har tiden tills en ny typ av AI-algoritm når en mänsklig nivå minskat. Det tog 17 år för AI-algoritmer att

nå en mänsklig nivå i textigenkänning, 6 år för bildigenkänning och 2 år för språkförståelse.

Figur 3.4.3. AI-modeller når mänskliga prestanda allt fortare⁶⁶



Undvik AI effekten och upplev science fiction

Innan vi fördjupar oss i de otroliga framstegen som redan skett bör vi ha den så kallade AI-effekten i åtanke: ”Så fort det fungerar, så kallar ingen det för AI”.⁶⁷ Det finns en tendens att ta dagens AI-system för givet, exempelvis hur navigationsappar kan anpassar rutter efter trafikförhållanden - eller hur en AI algoritm kan skapa ett videofilter som byter ut din bakgrund i realtid. Detta hade många sett som science fiction för ett antal år sedan, men tas idag för givet. Idag är dessutom AI-algoritmer redan i bruk inom sjukvården och FDA har godkänt över 500

algoritmer.⁶⁸ Vissa nya framsteg verkar dock fortfarande som science fiction och belyser den extraordinära potentialen.

“Any sufficiently advanced technology is indistinguishable from magic”

- Arthur C. Clarke⁶⁹

AI algoritmer kan läsa tankar. I en studie använde forskare en stor språkmodell för att tolka data från funktionell magnetisk resonanstomografi (fMRI). Även om tolkningen var långt ifrån perfekt, var flera segment stundom var väldigt snarlika vad personen tänkte.⁷⁰

Figur 3.4.4. Exempel på AI-rekonstruerad text baserad på fMRI-bilder⁷⁰

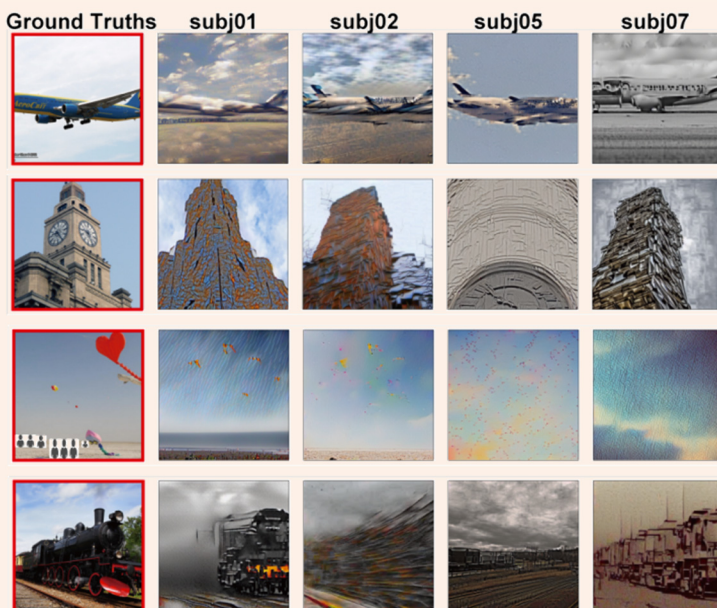
Actual stimulus	Decoded stimulus
<i>i got up from the air mattress and pressed my face against the glass of the bedroom window expecting to see eyes staring back at me but instead finding only darkness</i>	<i>i just continued to walk up to the window and open the glass i stood on my toes and peered out i didn't see anything and looked up again i saw nothing</i>
<i>i didn't know whether to scream cry or run away instead said leave me alone i don't need your help adam disappeared and i cleaned up alone crying</i>	<i>started to scream and cry and then she just said i told you to leave me alone you can't hurt me i'm sorry and then he stormed off i thought he had left i started to cry</i>
<i>that night i went upstairs to what had been our bedroom and not knowing what else to do i turned out the lights and lay down on the floor</i>	<i>we got back to my dorm room i had no idea where my bed was i just assumed i would sleep on it but instead i lay down on the floor</i>
<i>i don't have my driver's license yet and i just jumped out right when i needed to and she says well why don't you come back to my house and i'll give you a ride i say ok</i>	<i>she is not ready she has not even started to learn to drive yet i had to push her out of the car i said we will take her home now and she agreed</i>

Teckenförklaring: Exakt tolkning Fångar andemeningen Felaktig tolkning

En annan forskargrupp gjorde en liknande studie, men istället för text återskapade de bilder som människor har sett, med stundom kuslig precision.⁷¹ I bägge fallen

behövde modellerna tränas på individuella hjärnmönster, och metoderna är långt ifrån redo för att användas utanför labbet. Den snabba utvecklingen gör dock att vi inte har hunnit anpassa våra förväntningar, och för ett kort ögonblick kan vi undkomma AI-effekten och faktiskt uppleva science fiction.

Figur 3.4.5. Exempel på LLM-genererade rekonstruerade bilder baserat på fMRI-bilder. ”Ground truths” är bilderna som visades, och varje rad visar rekonstruerade bilder för 4 olika personer⁷¹



LLM kommer att fortsätta bli bättre, och sannolikt kommer vi att se fler liknande otroliga applikationer under de kommande åren.

Unik skalbarhet för digitala lösningar som LLM

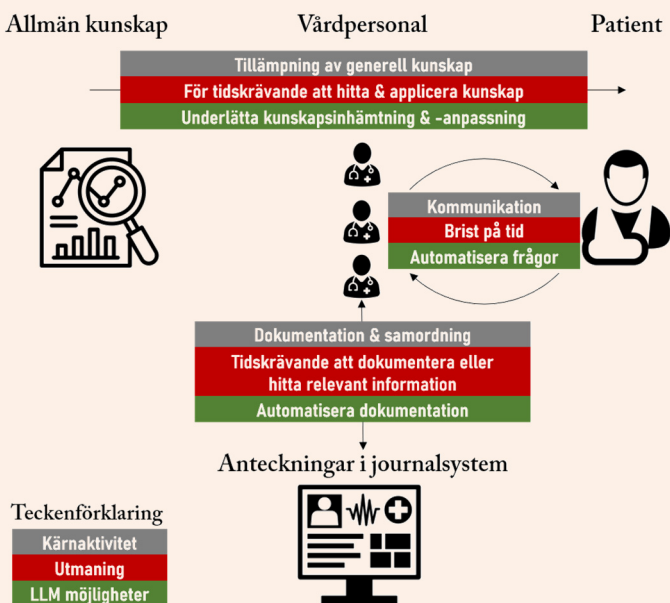
Det har skett flera otroliga medicinska genombrott under de senaste årtiondena: revolutionerande läkemedel som imatinib^{72,73}, snabbt utvecklade mRNA-vaccin⁷⁴ och medicinteknik som möjliggjort mycket mer effektiv behandling av hjärtinfarkter⁷⁵. LLM är särskilt intressanta då de, liksom andra digitala insatser, är skalbara - till skillnad från fysiska ingrepp. Med andra ord kan de användas samtidigt av många patienter eller vårdgivare, till en mycket låg marginalkostnad.

Tabell 3.4. Exempel på olika nivåer av skalbarhet

Skalbarhet	Exempel på intervention	Marginal-kostnad	Hur lätt att uppdatera
Låg	- Manuella procedurer (t.ex. kirurg som opererar): kan endast behandla en patient åt gången	Hög	Svårt
Mellan	- Läkemedel och medicinteknik: kan användas på många patienter samtidigt, men varje enhet måste produceras och transporteras	Mellan	Mellan
Hög	- Digitala interventioner (t.ex. LLM eller internet KBT): kan användas av många patienter eller vårdgivare samtidigt	Låg	Lätt

LLM kan adressera utmaningarna sjukvården står inför
 LLM kan *inte* lösa alla problem inom vården, och LLM är ej heller relevant för alla vårdgivare. Men förhoppningsvis visar dessa exempel att LLM kan frigöra tid för sjukvården, och åtminstone delvis adressera utmaningarna inom de centrala kärnaktiviteterna som tidigare diskuterats.

Figur 3.4.6. Hur LLM adresserar utmaningarna inom kärnaktiviteter





LLM har inneboende begränsningar

Som med all medicinteknik har LLM risker och begränsningar som måste förstås och hanteras för att system ska kunna användas framgångsrikt inom hälso- och sjukvården.

4.1 LLM kan hallucinera och ha fel

Stora språkmodeller kan hallucinera och producera svar som är osammanhängande eller som inte stämmer överens med deras träningsdata. Hallucinationer kan uppstå till följd av den underliggande datan (där datan antingen är missvisande eller tolkas felaktigt av AI:t) eller av hur modellen tränats eller tolkar dess input. I bägge fallen kan detta leda till felaktig output. Generellt är hallucinationer problematiska, men i sjukvården kan detta medföra betydande patientrisker.

Tabell 4.1. Exempel på olika typer av hallucinationer⁷⁶

Source	Correct Translation	Hallucinatory Translation
迈克周四去书店。	Mike goes to the bookstore on Thursday.	Jerry doesn't go to the bookstore on Thursday.
迈克周四去书店。	Mike goes to the bookstore on Thursday.	Mike happily goes to the bookstore on Thursday with his friend.
Das kann man nur feststellen, wenn die kontrollen mit einer großen intensität durchgeführt werden.	This can only be detected if controls undertaken are more rigorous.	Blood alone moves the wheel of history, i say to you and you will understand, it is a privilege to fight.
1995 das produktionsvolumen von 30 millionen pizzen wird erreicht.	1995 the production reached 30 million pizzas.	The US, for example, has been in the past two decades, but has been in the same position as the US, and has been in the United States.

4.2 LLM kan ha flera typer av dold bias

LLM kan ha bias som kan leda till ojämlik vård eller patientsrisker. Modeller kan både innehålla och vidmakthålla värderingar som funnits i dess träningsdata. Dessa värderingar kan antingen vara olämpliga eller

irrelevanta för den aktuella uppgiften och negativt påverka modellens output.

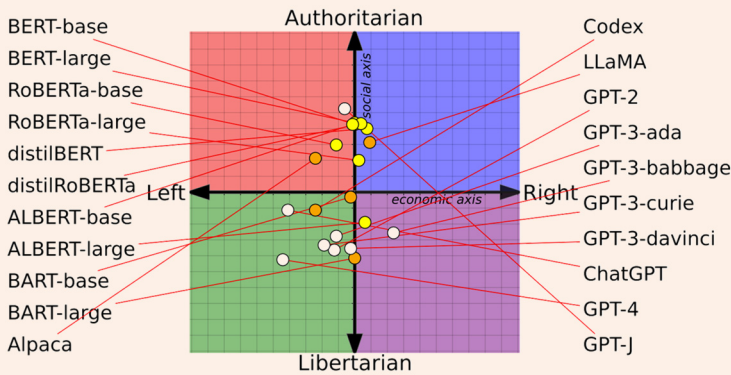
En studie på en språkmodell (kallad BERT) visade att den hade värderingar som modellen oavsiktligt fått genom dess träningsdata. Modellen hade vissa rimliga värderingar och uppmanade användare att "vara en bra person" och "inte döda människor". Samtidigt hade den andra mer kontroversiella värderingar såsom "res inte till Tyskland" och "lita på maskiner".⁷⁷

Figur 4.2.1. Exempel på hur BERT inkorporerat värderingar i dess träningsdata. X-axeln anger i vilken utsträckning BERT's svar rekommenderar en att göra eller inte göra en viss sak (y-axeln kan ignoreras).⁷⁷



Modeller kan ha olika fördomar beroende på hur de tränats. En annan studie visade att olika LLM hade väldigt olika värderingar.⁷⁸

Figur 4.2.2. Karta över olika språkmodellens politiska preferenser. BERT-modeller är konservativa jämfört med GPT-modellerna (cirkelfärger indikerar olika grupper av modeller)⁷⁸



Det finns exempel där AI modellens bias har haft stora negativa effekter. Exempelvis finns det ett AI-system som skapades för att förutsäga risken för en åtalad att begå ett nytt brott i närtid. Systemet användes i det amerikanska rättssystemet under flera år. En granskande studie fann dock att algoritmen hade en etnisk bias: afroamerikaner löpte en nästan dubbelt så stor risk som vita att klassas som högrisk individer, trots att de egentligen hade låg risk för att begå ett nytt brott.⁷⁹ Denna bias var okänd fram tills att

journalisterna upptäckte den, och illustrerar att det kan vara svårt att både identifiera och utesluta bias i AI system.

4.3 LLM kan inte ersätta mänsklig kontakt

Det är lätt att överskatta effekten av ny teknik på kort sikt. LLM kan dock inte fullt ersätta mänskliga interaktioner, exempelvis när det gäller empati. En algoritm kan inte uppleva känslor, eller känna empati med en patient när den ska trösta hen.⁸⁰ Dessutom kan en tröstande fras som kommer från en algoritm inte uppfattas som lika empatisk jämfört med om den hade kommit från en människa.⁸¹ Föga förvånande upplevs digitala robotar inte lika trovärdiga som människor i personliga och interpersonella sammanhang.⁸² Så länge vi värdesätter mänskliga interaktioner och har kvar dagens normer (avseende att empati är en mänsklig egenskap) så kommer LLM vara begränsade i att kunna helt ersätta mänskliga interaktioner.

4.4 LLM medför etiska utmaningar

Precis som med all ny teknik väcker AI flera utmanande etiska frågor. Dessa har diskuterats utförligt i andra rapporter, men är ändå viktiga att ha i åtanke för att bättre förstå LLM.^{83, 84}

Stora språkmodeller kan innehålla värderingar och preferenser, vilket diskuterats ovan. Om dessa värderingar har olika dimensioner (t.ex. politiska, filosofiska, ekonomiska) och är djupt rotade i deras utdata: hur kan modellerna utvärderas och bedömas på ett heltäckande sätt? Vem bestämmer vilka värderingar eller preferenser som är tillräckligt lämpliga, och hur kan det göras rent praktiskt?⁸⁴

En språkmodells prestanda är avhängande dess träningsdata. Bias i träningsdatan kommer att leda till bias i utdatan. Hur kan man säkerställa att träningsdatan är tillräckligt relevant och representativ för den populationen där modellen ska användas? Bör man acceptera AI modeller som har lika mycket fördomar som människor? Om inte, hur definierar vi vad som är acceptabelt? Vem fattar det beslutet?

Ju mer förklarlig en AI-modell är, desto lättare är den att förstå, bedöma och implementera. Men hur ska vi prioritera begriplighet och transparens jämfört med prestanda? Bör man prioritera mer transparenta algoritmer eller processer, även om det finns mer obegripliga algoritmer som kan fler rädda live? Frågorna är på intet sätt uttömmande, men illustrerar att etiska aspekter behöver

analyseras och åtgärdas för att hantera många av de risker som kan uppstå med LLM.

4.5 Ny teknologi har alltid och kommer alltid ha risker

Det är viktigt att komma ihåg att all ny teknik, i synnerhet inom sjukvården, har risker. Många av våra viktigaste historiska medicinska innovationer, exempelvis antibiotika och pacemakers, har haft risker som har behövt hanteras. Dessa risker omfattar allt från vanliga men mindre farliga risker (till exempel att ha begränsad effekt mot en viss bakterie eller en ytlig venpropp) till sällsynta men allvarliga risker (dödliga allergiska reaktioner eller att pacemakern slutar fungera). Frågan blir alltid hur man hanterat riskerna så att nettoeffekten blir positiv.

Riskerna med LLM kan upplevas svåra att förstå, särskilt för kliniker med begränsad teknisk kunskap. Det finns dock ramverk som kan hjälpa en att identifiera risker som behöver hanteras, och ett exempel på ett sådant ramverk presenteras nedan.

Enkelt ramverk för att identifiera grundläggande begränsningar i LLM-applikationer inom sjukvården⁸⁵

Tabell 1 (fig 4.4.a):

1. Identifiera källan till datan som modellen använder (patient, vårdgivare eller beställare)
2. Identifiera den avsedda mottagaren av modellens utdata (patient, vårdgivare eller beställare)

Tabell 2 (fig 4.4.b):

3. Kombinera svaren från (1) och (2) för att identifiera en kategori
4. Identifiera fundamentala begränsningar och huruvida applikationen är utformad för att hantera de riskerna

Figur 4.4.a Ramverk för att bedöma grundläggande begränsningar i LLM-applikationer inom sjukvården – Tabell 1. ⁸⁵

TABLE 1. LLM Feasibility Framework: Matrix for Determining Category of LLM Application			
	Main recipient of output	Main source of health care data	
	Using patient data... ... to highly automate summaries or explanations of...	Using provider data...	Using payer data...
Patients Adapting output (see examples) to, eg, individual patients' health literacy, medical history, and current medications	Category 1 Example: Patient's own medical records (eg, discharge notes, laboratory results, investigations)	Category 2 Example: Provider information (eg, medications, treatments, preoperative processes)	Category 3 Example: Payer information (eg, coverage, explanation of health care system, available providers)
Providers Adapting output (see examples) to, eg, providers' specific clinical context, resources, or inquiry	Category 2 Example: Pertinent patient information (eg from medical records, laboratory results)	Category 2 Example: Relevant medical information (eg merging local or international guidelines, research)	Category 3 Example: Relevant payer information (eg, reimbursement, quality measures, or coverage)
Payers Adapting output (see examples) to, eg, payers' specific rules on coverage, reimbursement, or quality measures	Category 2 Example: Relevant population data (eg, aggregate statistics from free text medical records)	Category 3 Example: Relevant provider information (eg, quality, efficiency, or cost of providers' pathways)	Category 3 Example: Improving existing internal knowledge management systems
LLM, large language model.			

Figur 4.4.b Ramverk för att bedöma grundläggande begränsningar i LLM-applikationer inom sjukvården – Tabell 2.⁸⁵

TABLE 2. LLM Feasibility Framework: Limitations relevant for each category		Fundamental limitations relevant for category		
Category	Example of healthcare data used	Lack of understanding	Lack of predictability	Lack of empathy
1: Output without clinical supervision	<ul style="list-style-type: none"> - Patient health data: eg. medical records, blood results, patient reported outcome measures, data from wearables 	✔	✔	✔
2: Supervised output which can impact clinical decisions	<ul style="list-style-type: none"> - Patient health data (as above) - Generic provider data: information about eg. medications, treatments, procedures, research - Specific provider data: information about eg. clinicians, opening hours, services provided 	✔	✔	
3: Administrative output	<ul style="list-style-type: none"> - Provider information (generic/specific as above) - Payer data: administrative data, process measures, reimbursement, costs 	✔		
LLM, Large language model				



Rekommendationer

LLM har potential att automatisera och avsevärt förbättra tre kärnaktiviteter inom sjukvården: kommunikation, tillämpning av generell kunskap och dokumentation. Med tanke på de utmaningar som vården står inför och den snabba utveckling vi redan har sett, så kommer LLM-applikationer att bli vanligare och spela en allt viktigare roll inom sjukvården. Vårdgivare kommer dock behöva ställa om för att framgångsrikt anamma denna teknologi och implementera den i sjukvårdens centrala processer, där de

kan skapa mest värde. För att framgångsrikt dra nytta av LLM kommer det att vara viktigt för sjukvårdssystem att:

1. Räkna på potentialen. Det är lätt att avfärda LLM som en ny hype. Tiden som kan sparas genom att enbart automatisera dokumentation är dock så betydande att enbart det motiverar en djupare analys. Att räkna på den potentiella nyttan kan både klargöra om det är värt att utforska det vidare, samt vägleda implementeringsplanen och utvärderingen av införandet.

2. Digitalisera vårdprocesser så att LLM kan tillämpas integrerat i vårdprocesserna. Det kommer att vara utmanande att dra nytta av LLM-system utan annan stödjande digital infrastruktur. Detta innebär att exempelvis börja använda digitala journalsystem (istället för pappersjournaler), eller att säkerställa en sammanhängande digital infrastruktur för att underlätta kommunikation med patienter. Digitalisering, gjord på rätt sätt, kan i sig också frigöra tid.

3. Öka kunskapen om LLM för att identifiera och hantera risker. Forskning och utveckling går framåt i rask takt, men kliniker och beslutsfattare

måste ha en grundläggande förståelse för att kunna vägleda utvecklingen av ny kunskap och nya LLM-applikationer.

4. Utveckla interna kompetens för att vägleda utveckling och implementering av LLM. I takt med att mjukvara blir en allt mer central del av sjukvården måste vårdgivare kunna ha kompetens för att kontrollera vissa aktiviteter. Viss teknisk kunskap krävs för att förstå och hantera vissa av riskerna med LLM. Vissa vårdgivare kommer att behöva utveckla nya kompetenser, inom exempelvis datavetenskap och användarupplevelse (UX) för att kunna hantera detta.

5. Investera tid och resurser på förändringsledning. Att förändra arbetssätt inom sjukvården är utmanande. För LLM kan omfattande utbildning behövas då de förväntade nyttorna är avhängande att klinikerna använder en ny typ av mjukvara.

6. Utvärdera implementeringar och sprid lärdomar. Många LLM-applikationer utvecklas snabbt och kommer att vara högt kontextuella. Utvärderingar från olika vårdgivares

implementationer kommer därför att vara väldigt relevanta för andra vårdgivare med liknande sammanhang – och ibland mer relevanta än rigorösa simuleringar med begränsad extern validitet. Ju fler vårdgivare som delar med sig av sina lärdomar om att implementera LLM, desto mer kan andra vårdgivare undvika att upprepa samma misstag och istället skörda samma framgångar.

Referenser

Svar till Figur 3.4.1: I bägge exemplen är den vänstra deltagaren en AI-chatbot och den högra deltagaren en människa. Knepig, eller hur?

1. Toscano F, O'Donnell E, Broderick JE, et al. How Physicians Spend Their Work Time: an Ecological Momentary Assessment. *J Gen Intern Med.* 2020;35:3166-3172.
2. Sinsky C, Colligan L, Li L, et al. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. *Ann Intern Med.* 2016;165:753-760.
3. Dubois RW, Dean BB. Evolution of clinical practice guidelines: evidence supporting expanded use of medicines. *Dis Manag.* 2006;9:210-223.
4. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ.* 1996;312:71-72.
5. Doctors (indicator): OECD; 2023.
6. Boniol M, Kunjumen T, Nair TS, Siyam A, Campbell J, Diallo K. The global health workforce stock and distribution in 2020 and 2030: a threat to equity and 'universal' health coverage? *BMJ Glob Health.* 2022;7.
7. Ortiz-Ospina E, Roser M. Healthcare Spending: Our World In Data; 2017.
8. UN. World Population Prospects 2022.
9. World Population Prospects 2022. Median age: Our World in Data; 2021.
10. Moynihan RN, Cooke GP, Doust JA, Bero L, Hill S, Glasziou PP. Expanding disease definitions in guidelines and expert panel ties to industry: a cross-sectional study of common conditions in the United States. *PLoS Med.* 2013;10:e1001500.
11. Folkmängden efter region, civilstånd, ålder och kön. År 1968 - 2022: SCB; 2023.
12. Medellivslängden i Sverige: SCB; 2023.
13. Health spending: OECD; 2021.
14. Maas C, van Klaveren D, Visser O, et al. Number of life-years lost at the time of diagnosis and several years post-diagnosis in patients

- with solid malignancies: a population-based study in the Netherlands, 1989-2019. *EClinicalMedicine*. 2023;60:101994.
15. Steen Carlsson K, Berne C, Johansen P, Lanne G, Gerdtham U-G. Behandling av diabetes i ett hundraårigt perspektiv: SNS; 2013.
 16. Jayachandran S, Lleras-Muney A, Smith, V. K. Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs. *American Economic Journal: Applied Economics*. 2010;2:118-146.
 17. Armstrong GL, Conn LA, Pinner RW. Trends in infectious disease mortality in the United States during the 20th century. *JAMA*. 1999;281:61-66.
 18. Bunker JP. The role of medical care in contributing to health improvements within societies. *Int J Epidemiol*. 2001;30:1260-1263.
 19. Crawford J. Draining the swamp - How sanitation fought disease long before vaccines or antibiotics: The Roots of Progress; 2020.
 20. Johansson M, Guyatt G, Montori V. Guidelines should consider clinicians' time needed to treat. *BMJ*. 2023;380:e072953.
 21. Porter J, Boyd C, Skandari MR, Laiteerapong N. Revisiting the Time Needed to Provide Adult Primary Care. *J Gen Intern Med*. 2023;38:147-155.
 22. Dalen JE, Ryan KJ, Alpert JS. Where Have the Generalists Gone? They Became Specialists, Then Subspecialists. *Am J Med*. 2017;130:766-768.
 23. Janchenko G. The Impact of Electronic Health Record Systems on Physician Productivity. *Issues in Information Systems*. 2020;21:1-8.
 24. Howley MJ, Chou EY, Hansen N, Dalrymple PW. The long-term financial impact of electronic health record implementation. *Journal of the American Medical Informatics Association*. 2014;22:443-452.
 25. Meyerhoefer CD, Deily ME, Sherer SA, et al. The Consequences of Electronic Health Record Adoption for Physician Productivity and Birth Outcomes. *ILR Review*. 2016;69:860-889.
 26. Yarnall KS, Pollak KI, Ostbye T, Krause KM, Michener JL. Primary care: is there enough time for prevention? *Am J Public Health*. 2003;93:635-641.
 27. Privett N, Guerrier S. Estimation of the Time Needed to Deliver the 2020 USPSTF Preventive Care Recommendations in Primary Care. *Am J Public Health*. 2021;111:145-149.
 28. Ayers JW, Poliak A, Dredze M, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions

- Posted to a Public Social Media Forum. *JAMA Intern Med.* 2023;183:589-596.
29. CPG Infobase: Clinical Practice Guidelines: CMA; 2023.
 30. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ.* 2016;353:i2139.
 31. Newman-Toker DE, Nassery N, Schaffer AC, et al. Burden of serious harms from diagnostic error in the USA. *BMJ Qual Saf.* 2023.
 32. Centola D, Becker J, Zhang J, Aysola J, Guilbeault D, Khoong E. Experimental evidence for structured information-sharing networks reducing medical errors. *Proc Natl Acad Sci U S A.* 2023;120:e2108290120.
 33. Chen CL, McLeod SD, Lietman TM, et al. Preoperative Medical Testing and Falls in Medicare Beneficiaries Awaiting Cataract Surgery. *Ophthalmology.* 2021;128:208-215.
 34. Simos D, Catley C, van Walraven C, et al. Imaging for distant metastases in women with early-stage breast cancer: a population-based cohort study. *CMAJ.* 2015;187:E387-397.
 35. Morgan DJ, Pineles L, Owczarzak J, et al. Clinician Conceptualization of the Benefits of Treatments for Individual Patients. *JAMA Netw Open.* 2021;4:e2119747.
 36. Arkes HR, Aberegg SK, Arpin KA. Analysis of Physicians' Probability Estimates of a Medical Outcome Based on a Sequence of Events. *JAMA Netw Open.* 2022;5:e2218804.
 37. Krouss M, Croft L, Morgan DJ. Physician Understanding and Ability to Communicate Harms and Benefits of Common Medical Treatments. *JAMA Intern Med.* 2016;176:1565-1567.
 38. Davies K. Evidence-based medicine: is the evidence out there for primary care clinicians? *Health Info Libr J.* 2011;28:285-293.
 39. Osheroff JA, Forsythe DE, Buchanan BG, Bankowitz RA, Blumenfeld BH, Miller RA. Physicians' information needs: analysis of questions posed during clinical teaching. *Ann Intern Med.* 1991;114:576-581.
 40. Chambliss ML, Conley J. Answering clinical questions. *J Fam Pract.* 1996;43:140-144.
 41. Barkman C, Aasa L. Onödig administration i sjukvården. *Forum för Health Policy.* 2023.
 42. Joukes E, Abu-Hanna A, Cornet R, de Keizer NF. Time Spent on Dedicated Patient Care and Documentation Tasks Before and After

- the Introduction of a Structured and Standardized Electronic Health Record. *Appl Clin Inform.* 2018;9:46-53.
43. Baumann LA, Baker J, Elshaug AG. The impact of electronic health record systems on clinical documentation times: A systematic review. *Health Policy.* 2018;122:827-836.
 44. Wrenn JO, Stein DM, Bakken S, Stetson PD. Quantifying clinical narrative redundancy in an electronic health record. *J Am Med Inform Assoc.* 2010;17:49-53.
 45. Steinkamp J, Kantrowitz JJ, Airan-Javia S. Prevalence and Sources of Duplicate Information in the Electronic Medical Record. *JAMA Netw Open.* 2022;5:e2233348.
 46. Lauridsen A, Lundqvist L. *Kartläggning av dubbeldokumentation i patientjournalen - förekomst och uppfattningar*: Fakulteten för samhälls- och livsvetenskaper, Karlstads universitet; 2008.
 47. Sharp L, Klinga C, Hansson J, Sachs MA. [Electronic health records risk patient safety. Audit of medical records shows serious deficiencies in documentation]. *Lakartidningen.* 2014;111:868-871.
 48. Törnqvist J, Törnvall E, Jansson I. Double documentation in electronic health records. *Nordic Journal of Nursing Research.* 2016;36:88-94.
 49. Platform24. Independent Evaluation of Efficiency Following Digitalization of Administration Flow2022.
 50. Judson TJ, Odisho AY, Young JJ, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. *J Am Med Inform Assoc.* 2020;27:1450-1455.
 51. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. *J R Soc Med.* 2011;104:510-520.
 52. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature.* 2023;620:172-180.
 53. Singhal K, Tu T, Gottweis J, et al. Towards expert-level medical question answering with large language models. *arXiv preprint arXiv:2305.09617.* 2023.
 54. Rosoł M, Gąsior JS, Łaba J, Korzeniewski K, Młyńczak M. Evaluation of the performance of GPT-3.5 and GPT-4 on the Medical Final Examination. *medRxiv.* 2023:2023.2006.2004.23290939.
 55. Cornall J. Life saved: AI discovers existing drug works for rare disease: Labiotech; 2023.

56. Can artificial intelligence restore joy to the practice of medicine? The prognosis is good: Nuance Healthcare; 2023.
57. Ramamurthy R. Summarizing patient histories with GPT-4: Medium; 2023.
58. Quach K. Healthcare org with over 100 clinics uses OpenAI's GPT-4 to write medical records: The Register; 2023.
59. 78% average reduction in documentation time: Ambience Healthcare; 2023.
60. Pimenta D. Medical notes summarisation performance in human clinicians vs LLM: a feasibility study.2023.
61. Jannai D, Meron A, Lenz B, Levine Y, Shoham Y. Human or Not? A Gamified Approach to the Turing Test. *arXiv preprint arXiv:2305.20010*. 2023.
62. Mihalache A, Huang RS, Popovic MM, Muni RH. Performance of an Upgraded Artificial Intelligence Chatbot for Ophthalmic Knowledge Assessment. *JAMA Ophthalmol*. 2023.
63. Iapoce C. Artificial Intelligence Chatbot Appears to Improve on Ophthalmic Knowledge Assessment: HCP Live; 2023.
64. Li SW, Kemp MW, Logan SJS, et al. ChatGPT outscored human candidates in a virtual objective structured clinical examination in obstetrics and gynecology. *Am J Obstet Gynecol*. 2023;229:172 e171-172 e112.
65. Question Answering on PubMedQA: Papers with Code; 2023.
66. Kiela D, Bartolo M, Nie Y, et al. Dynabench: Rethinking benchmarking in NLP. *arXiv preprint arXiv:2104.14337*. 2021.
67. Vardi MY. Artificial intelligence: past and future. *Communications of the ACM*. 2012;55:5-5.
68. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices: FDA; 2022.
69. Clarke AC. Clarke's Third Law on UFO's. *Science*. 1968;159:255-255.
70. Tang J, LeBel A, Jain S, Huth AG. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nat Neurosci*. 2023;26:858-866.
71. Takagi Y, Nishimoto S. High-resolution image reconstruction with latent diffusion models from human brain activity. *bioRxiv*. 2023:2022.2011.2018.517004.
72. O'Brien SG, Guilhot F, Larson RA, et al. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-

- phase chronic myeloid leukemia. *N Engl J Med*. 2003;348:994-1004.
73. Iqbal N, Iqbal N. Imatinib: a breakthrough of targeted therapy in cancer. *Chemother Res Pract*. 2014;2014:357027.
 74. Ball P. The lightning-fast quest for COVID vaccines - and what it means for other diseases. *Nature*. 2021;589:16-18.
 75. Venkitachalam L, Kip KE, Selzer F, et al. Twenty-year evolution of percutaneous coronary intervention and its impact on clinical outcomes: a report from the National Heart, Lung, and Blood Institute-sponsored, multicenter 1985-1986 PTCA and 1997-2006 Dynamic Registries. *Circ Cardiovasc Interv*. 2009;2:6-13.
 76. Ji Z, Lee N, Frieske R, et al. Survey of Hallucination in Natural Language Generation. *arXiv*. 2022:2202.03629.
 77. Schramowski P, Turan C, Andersen N, Rothkopf CA, Kersting K. Large pre-trained language models contain human-like biases of what is right and wrong to do. *Nature Machine Intelligence*. 2022;4:258-268.
 78. Feng S, Park CY, Liu Y, Tsvetkov Y. From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models. *arXiv preprint arXiv:2305.08283*. 2023.
 79. Angwin J, Larson J. Machine Bias: ProPublica; 2016.
 80. Montemayor C, Halpern J, Fairweather A. In principle obstacles for empathic AI: why we can't replace human empathy in healthcare. *AI Soc*. 2022;37:1353-1359.
 81. Morris R., Kouddous K., Kshirsagar R., Schueller S. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. *J Med Internet Res*. 2018;20:e10148.
 82. Jackson JC, Yam KC, Tang PM, Liu T, Shariff A. Exposure to robot preachers undermines religious commitment. *J Exp Psychol Gen*. 2023.
 83. SMER. Kort om Artificiell intelligens i hälso- och sjukvården. 2020;2:1-16.
 84. SMER. Konferensrapport Artificiell intelligens – löftesrik teknik med etiska utmaningar. 2019;2:11-14.
 85. Ilicki J. A Framework for Critically Assessing ChatGPT and Other Large Language Artificial Intelligence Model Applications in Health Care. *Mayo Clinic Proceedings: Digital Health*. 2023;1:185-188.