
Kan vi använda AI för att skapa nya riktlinjer och bedöma vetenskaplig evidens inom hjärt-lungräddning?

MARKUS SKRIFVARS

Artificiell intelligens (AI) är högaktuellt inom medicinen och ett flertal olika användningsområden har diskuterats. En av AI:s egenskaper är att den kan gå igenom stora mängder data snabbt. Eftersom antalet vetenskapliga publikationer ökar nästan exponentiellt, har genomgång av befintlig vetenskaplig litteratur för att utarbeta och sammanställa behandlingsriktlinjer föreslagits som ett område där AI kunde vara till hjälp. Inom hjärt-lungräddning (HLR) har användning av så kallad hypotermibehandling som metod för att lindra en hjärnskada som uppstår efter ett hjärtstillestånd diskuterats flitigt. Ett flertal studier har genomförts med varierande resultat. Fyra stora europeiska organisationer inom akut- och intensivvård har nämligen kommit fram till helt olika behandlingsrekommendationer. En forskningsgrupp har därför beslutat använda AI för att försöka utreda om hypotermibehandling är effektiv eller inte och för att se om AI kan vara till hjälp med att sammanfatta studierna och ge svar på varför organisationerna har tolkat den medicinska litteraturen så olika. Forskarna har använt sig av modellerna Chat GPT och Gemini. Resultatet blev att modellerna sammanfattade det som publicerats om effekten av hypotermibehandling. Men ingendera modellen gav någon mer kritisk granskning av vad som fick organisationerna att tolka litteraturen så olika, trots att detta har debatterats flitigt i flera vetenskapliga tidskrifter. Slutresultatet blev att AI inte var till någon egentlig hjälp för dem som behärskar ämnet. För dem som inte behärskar ämnet kan AI-modellernas analyser leda till direkt felaktiga slutsatser. Av någon anledning verkade båda modellerna ge de två äldsta studierna, som rapporterade positiva effekter av hypotermibehandling, mer vikt trots att kvaliteten på dem var mycket sämre än i flera nyare studier med motsatta resultat.

Är bilder skapade med artificiell intelligens bättre än en verklig bild tagen med kamera?

Alla talar numera om artificiell intelligens (AI). En ny AI-metod kan skapa bilder, filmsnuttar och musik på nolltid. Vissa av dem är så bra att det är svårt att skilja mellan det som skapats av AI och det ”riktiga”. Utan tvekan häpnadsväckande sett ur teknisk synvinkel! Visst är det spännande att nya bilder, filmer och annat kan skapas på nolltid, men för mig kvarstår trots allt grundfrågan ”Varför?”. Det är klart att all innovation leder till nya fynd och tillämpningsmöjligheter som vi inte har kunnat förutspå, men varför är en bild skapad av AI bättre än en kamerabild, som i dagens läge går lika snabbt att ta?

SKRIBENTEN

Markus Skrifvars, professor, med.o.kir.dr, EDICM, FCICM

Clinicum, Helsingfors universitet, och Kliniken för anestesiology och intensivvård, HUS

Användning av AI inom medicinen för diagnostik och analys av stora patientkohorter

Användningen av AI är utan tvekan ett av de mest hajpade hjälpmedlen inom medicinen för tillfället. Vi får nästan dagligen höra hur artificiell intelligens har överträffat läkare och ställt bättre diagnoser inom ett flertal

medicinska discipliner, inbegripet analys av röntgenbilder och avklarande av vissa tentamina (1–3). Rubriker som "Artificiell intelligens kommer att transformera medicinen" förekommer nästan dagligen och alla kongresser har en eller flera sessioner med AI som tema. Artificiell intelligens föreslås också som analysmetod för stora patientmaterial, så kallade *big data* (4). Men frågan är vad dessa analyser egentligen är och i vilka frågeställningar AI skulle överträffa alldeles vanliga statistiska flervariabelmodeller? Av någon orsak ställer jag mig lite skeptisk till allt detta. Man drar sig till minnes 1990- och 2000-talet när genetiska analyser var den stora förändring som skulle revolutionera medicinen. Man såg för sig en framtid, där patienten kommer till sjukhuset, det tas ett blodprov och görs en analys av patientens gener, varefter det planeras individuell behandling utifrån patientens gener. Detta har säkert hänt inom vissa specialiteter, som i vissa fall inom onkologin, men inom vanlig akutmedicin och intensivvårdsmedicin har detta inte hänt, och frågan är om det någonsin kommer att hända. De flesta vanliga sjukdomar är multifaktoriella – och levnadsvanorna spelar en mycket stor roll. Samtidigt är de möjliga behandlingsmetoderna i grund och botten ganska begränsade för många vanliga sjukdomar, som lunginflammation, sepsis, hjärtinfarkt och stroke. Kanske härrör min skepsis från att trenden, speciellt inom intensivvård, under de senaste åren alldeles tydligt har varit att "*less is more*", det vill säga mindre medicinering, mindre nedsövning, mindre syre, mindre användning av medicinsk utrustning. Och de stora framstegen vad beträffar specifika nya läkemedel har helt uteblivit (5).

AI som hjälpmedel för vetenskaplig evaluering av behandlingsmetoder

Naturligtvis kommer AI att förändra mycket av det vi gör, men hur är en annan fråga. En fördel med AI är självfallet dess förmåga att snabbt gå igenom data, exempelvis nya publikationer inom ett visst område. Inom hjärt-lungräddning (HLR) publiceras inter-

nationella behandlingsriktlinjer vart femte år av organisationen *International Liaison Committee on Resuscitation (ILCOR)*. Organisationen är indelad i arbetsgrupper som behandlar olika områden av HLR. Jag har själv arbetat inom ILCOR i femton år och bidragit till arbetet med att vetenskapligt evaluera olika behandlingsmetoder inom HLR. Utifrån dessa sammanfattningar kan sedan andra internationella vetenskapliga organisationer, som *European Society of Intensive Care* och *European Resuscitation Council* och även Duodecims arbetsgrupp för god medicinsk praxis sammanställa praktiska riktlinjer för vården och behandlingen (6). Överlag är denna arbetsprocess mycket tidskrävande och en högaktuell fråga är om AI kan snabba upp processen.

Hypotermibehandling och dess effektivitet efter hjärtstillestånd

Ett mycket omdebatterat område inom vården av patienter efter HLR är hypotermibehandling, som innebär nedkylning av patientens kroppstemperatur till 33 °C i 24 timmar. Det finns en del experimentell forskning som visar på positiva effekter av nedkylning, och vid nedkylning som sker innan blodcirkulationen avstannar har en skyddande effekt kunnat bevisas. Denna metod används inom vissa typer av krävande hjärtkirurgi. Men huruvida en sänkning av temperaturen *efter* att patienten haft ett hjärtstillestånd kan minska omfattningen av organskada är en helt annan sak. År 2002 publicerades två studier som påvisade en positiv effekt av hypotermibehandling hos patienter efter hjärtstillestånd (7, 8) och de kom att fullständigt förändra behandlingen av dessa patienter. Medicinsk apparatur utvecklades för att snabbare och enklare introducera metoder att upprätthålla låg temperatur hos patienter. Dessa två studier skulle i dag anses ha något sämre vetenskaplig kvalitet och även vara ganska små, vilket innebär att den effekt man då såg i grupperna tenderar att vara realistisk utifrån den information som man har idag. I och med två mycket större studier som

HUVUDBUDSKAP:

- Artificiell intelligens (AI) är högaktuellt inom medicinen. AI kan gå igenom stora mängder data på kort tid, men enligt ett exempel inom intensivvården var AI:s slutsatser om effektiviteten hos en viss vårdmetod direkt felaktiga. Trots att det skrivits om detta inom den vetenskapliga litteraturen nämnde inte AI det mest relevanta problemet inom området.

Can you draw a picture of Markus Skrifvars at work in the hospital.



Here is an artistic depiction of Markus Skrifvars at work in a hospital setting, showcasing him in his role as a professor of emergency medicine. The image portrays him in an intensive care unit, surrounded by medical equipment and a team, emphasizing his dedication to critical patient care.

Figur 1. Bild på Markus Skrifvars enligt AI (till vänster) och med kamera (till höger). Det är upp till läsaren att bedöma vilken bild som är "bäst". AI-bilden är producerad med Chat GPT i oktober 2024.

publicerades 2013 respektive 2021 kunde en internationell forskargrupp med två stora randomiserade multicenterstudier påvisa att hypotermibehandling som påbörjas på intensivvårdsavdelning efter HLR inte förbättrar vare sig överlevnad eller andelen patienter med god neurologisk funktionsnivå sex månader efter hjärtstillestånd (9, 10). Resultatet blev rentav att patienter som genomgått hypotermibehandling fick flera biverkningar och att överlevnaden var två procent lägre än hos patienter som behandlades endast om de fick feber (temperatur över 37,8 °C). Dessa två studier har lett till att vissa organisationer har omtolkat behandlingsrekommendationerna. I dagens läge rekommenderar ERC och ESICM att hypotermibehandling inte används på patienter efter HLR, medan andra organisationer som *European Society of Anaesthesiology and Intensive Care (ESAIC)* och *European Society of Emergency Medicine (EUSEM)* fortfarande förespråkar hypotermibehandling som standardbehandling efter HLR (6, 11). Saken blir ännu mer komplicerad av att två systematiska översiktsartiklar, den ena av gruppen Cochrane och den andra av ILCOR, har kommit fram till helt olika slutsatser trots att de utgått från samma studier (12, 13). Hur ska en kliniker tolka dessa olika rekommendationer?

AI som verktyg för att evaluera huruvida hypotermibehandling fungerar

En grupp forskare beslutade använda AI för att belysa varför tolkningarna av de vetenskapliga studierna är så olika (14). Forskarna använde sig av Chat GPT och Gemini och bad dem sammanfatta vad vi vet om effekten av hypotermibehandling på överlevnaden efter HLR. Båda modellerna producerade en relativt utförlig sammanställning över vad tidigare vetenskapliga studier har påvisat. Men modellerna kunde inte förklara vad som gjort att resultaten hade tolkats så olika, trots att detta har diskuterats flitigt i vetenskapliga tidskrifter (15). Alltså förblev de riktigt svåra och väsentliga frågorna obesvarade av båda AI-modellerna. Det tråkiga för mig som är insatt i ämnet är att AI i detta fall ger en del sammanfattande information som kan anses nyttig och användbar, men som inte ger svar på de viktiga frågorna eller kommer med felaktiga svar och slutsatser. Om någon som inte är insatt i ämnet skulle utarbeta behandlingsriktlinjer utifrån en sådan AI-analys vore det fullt möjligt att personen skulle dra direkt felaktiga slutsatser.

Analysprocessen för modellerna bör vara mera transparent och basera sig på a priori-planerade orsakssamband

Lite besvärad av detta har jag själv ett flertal

gångar ställt Chat GPT den enkla frågan "Förbättrar hypotermibehandling patienternas överlevnad och neurologiska funktionsnivå efter hjärtstillestånd?". Svaren har genomgående börjat med "enligt studier från 2002 förbättrar hypotermibehandling prognosen hos patienter efter HLR". Detta är enligt min mening oroväckande och det vore intressant att veta hur modellen resonerar i detta fall. Jag antar att det kan ha att göra med att den studie som publicerades 2002 kanske har gett större avtryck på internet och därför klassas som viktigare än de studier som gjordes 2013 och 2021. Detta är kanske ett exempel på fenomenet "*black box*", alltså att vi inte vet varför modellen kommer fram till den slutsats den gör. Ett liknande exempel har diskuterats när det gäller bedömning av personers kreditförmåga och möjligheten att få lån i en bank. En viktig etisk fråga är om vi är villiga att inkludera egenskaper som ras, kön, sexualitet och liknande i modeller av denna typ även om de statistiskt sett gör modellen bättre. En viktig princip inom medicinen har varit att undvika att analysera material för mycket utan att ha en förutbestämd plan för att minska risken för slumpmässiga fynd. Begreppen "*fishing expedition*" och "*I tortured the data until it confessed*" är relevanta i detta sammanhang och enligt min mening högrelevanta även med tanke på bruket av AI på massiva patientmaterial. Ett praktexempel från intensivvårdens värld är ett fynd från en studie om vätskebehandling av patienter med sepsis. I en kompletterande analys kunde forskarna visa att de patienter som var födda i stjärntecknet Fiskarna hade en signifikant lägre dödlighet om de behandlades med kolloid lösning än om de behandlades med kristalloid lösning, trots att den stora studien var helt neutral (16). Eftersom det inte finns någon rimlig patofysiologisk förklaring till detta, drar de flesta sannolikt slutsatsen att det är ett slumpmässigt fynd. Frågan är om vi har denna möjlighet med komplexa AI-modeller, där vi inte exakt vet hur modellen har kommit fram till sin slutsats. Ska vi inkludera variabler som verkar helt otroliga ur en patofysiologisk synvinkel?

Kan AI ersätta randomiserade kliniska studier i framtiden?

Randomiserade studier för att utvärdera om en viss behandlingsmetod fungerar eller inte är både dyra och tidskrävande. Även här har man sett AI som en lösning. Med AI-analys

av stora material kommer vi att kunna evaluera olika behandlingsmetoder snabbare och billigare. Personligen ställer jag mig mycket skeptisk till detta. Vi har ett flertal exempel från medicinen där observationsstudier med statistiska metoder kommit fram till att en viss behandlingsmetod, exempelvis adrenalin vid HLR och användning av intraosseös nål under HLR, är kopplad till högre dödlighet (17). Detta trots olika statistiska analysmetoder och stora patientmaterial. När sedan stora randomiserade studier har gjorts har dessa fynd förkastats totalt. Det beror på "*confounding by indication*". En kliniker använder eller låter bli att använda en metod på basis av en helhetsbedömning som inte kan beskrivas med de numeriska data som finns i patientjournaler. Till exempel patienter med en allmänt god prognos har kunnat återfå sin blodcirkulation innan man har kunnat ge adrenalin och överlever därför, men orsaken är inte att de inte har fått adrenalin, utan att hjärtat har startat innan de hunnit få det.

Problem av denna typ kommer att vara mycket svåra att undvika med de numeriska patientdata i form av vitala funktioner och blodprover som vi samlar i våra patientjournaler. Och då finns risken att en AI-modell kommer fram till felaktiga slutsatser. Om slutsatsen gäller en viss behandlingsmetod kommer vi inte att veta om slutsatsen är rätt eller fel, innan det har gjorts en stor randomiserad studie.

Sammanfattningsvis är det väl klart att AI kommer att förändra medicinen. Men på vilket sätt är en helt annan fråga som vi inte vet svaret på idag! I princip anser jag att vi måste se AI-modeller som till exempel stöd för behandlingsbeslut på samma sätt som nya läkemedel eller medicintekniska apparater. Det måste finnas tillförlitliga bevis på att de faktiskt fungerar, att de minskar kostnaderna eller att de förbättrar behandlingen och räddar liv. Om så inte är fallet, ställer jag mig skeptisk till att använda dessa modeller som underlag för medicinska beslut, särskilt när det gäller effekten av medicinska ingrepp och läkemedel. Personligen tror jag att AI kommer att bli ett hjälpmedel och inte så mycket annat. Visst måste man se med en viss skepsis på techbolag som prackar sina metoder och program också på oss inom medicinen.

Markus Skrifvars enligt Chat GPT

Om jag till sist återgår till samma slags navelskåderi som i inledningen och frågar AI

Markus Skrifvars is a Finnish-Swedish media personality known for his work as a television and radio host in Finland. He is part of the Swedish-speaking community in Finland and has gained recognition for his engaging style and presence on various media platforms.

Skrifvars has hosted several programs, often in the context of cultural and entertainment shows. His work primarily focuses on Finnish-Swedish media, catering to the Swedish-speaking population in Finland.

If you need more specific information about his career or personal life, please let me know!

🗨️ 📄 📌 🗑️ 🔍

Figur 2. Chat GPT svarar på frågan "Vem är Markus Skrifvars?" i oktober 2024.

"Vem är Markus Skrifvars?" får jag det svar (oktober 2024) som visas i figur 2.

Alltså ett totalt felaktigt svar. I sig är det intressanta *inte* att Chat GPT inte vet vem jag är. Jag anser inte att just jag är så enastående att Chat GPT borde veta det, utan det intressanta är att modellen verkar vara rätt så säker på sin sak fastän svaret är fel. Är det så vi vill ha det inom medicinen? Starka åsikter som är felaktiga? Vore det inte bättre att modellen ärligt svarade att det finns flera möjliga svar och att modellen är osäker? Jag ställde en del följdfrågor och slutligen svarade Chat GPT rätt. Glädjande nog (åtminstone för mitt ego) svarade Chat GPT helt rätt på samma fråga när jag ställde den i juni 2025. Modellen hade tydligen lärt sig av mina instruktioner och vet nu vem jag är (sic!). Så kanske det finns hopp? Åtminstone när det gäller den kanske inte för alla lika relevanta frågan "Vem är Markus Skrifvars?". Men det hade givetvis funnits andra sätt att komma till samma riktiga svar även utan AI. Men det kanske är irrelevant i sammanhanget.

Markus Skrifvars
markus.skrifvars@hus.fi

Inga bindningar

Referenser

- Wang D, Jin R, Shieh CC, Ng AY, Pham H, Dugal T, Barnett M, Winoto L, Wang C, Barnett Y. Real world validation of an AI-based CT hemorrhage detection tool. *Front Neurol*. 2023 Aug 3;14:1177723. doi: 10.3389/fneur.2023.1177723. PMID: 37602253; PMCID: PMC10435741.
- Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D. How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med Educ*. 2023; 9:e45312.
- Kuo KM, Chang CS. A meta-analysis of the diagnostic test accuracy of artificial intelligence predicting emergency department dispositions. *BMC Med Inform Decis Mak*. 2025 May 15;25(1):187. doi: 10.1186/s12911-025-03010-x. PMID: 40375078; PMCID: PMC1208289
- Tsai ML, Chen KF, Chen PC. Harnessing Electronic Health Records and Artificial Intelligence for Enhanced Cardiovascular Risk Prediction: A Comprehensive Review. *J Am Heart Assoc*. 2025 Mar 18;14(6):e036946. doi: 10.1161/JAHA.124.036946. Epub 2025 Mar 13. PMID: 40079336; PMCID: PMC12132661.
- Perner, A., Finfer, S. Do trials that report a neutral or negative treatment effect improve the care of critically ill patients? *Yes*. *Intensive Care Med* 44, 1985–1988 (2018). <https://doi.org/10.1007/s00134-018-5129-5>.
- Nolan JP, Sandroni C, Böttiger BW, Cariou A, Cronberg T, Friberg H, Genbrugge C, Haywood K, Lilja G, Moulart VRM, Nikolaou N, Mariero Olasveengen T, Skrifvars MB, Taccone F, Soar J. European Resuscitation Council and European Society of Intensive Care Medicine Guidelines 2021: Post-resuscitation care. *Resuscitation*. 2021 Apr;161:220-269. doi: 10.1016/j.resuscitation.2021.02.012. Epub 2021 Mar 24. Erratum in: *Resuscitation*. 2021 Oct;167:109-110. doi: 10.1016/j.resuscitation.2021.08.013. PMID: 33773827.
- Hypothermia after Cardiac Arrest Study Group. Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. *New England Journal of Medicine* 2002;346(8):549-56.
- Bernard SA, Gray TW, Buist MD, Jones BM, Silvester W, Gutteridge G, et al. Treatment of comatose survivors of out-of-hospital cardiac arrest with induced hypothermia. *New England Journal of Medicine* 2002;346(8):557-63.
- Nielsen N, Wetterslev J, Cronberg T, Erlinge D, Gasche Y, Hassager C, et al. TTM Trial Investigators. Targeted temperature management at 33°C versus 36°C after cardiac arrest. *New England Journal of Medicine* 2013;369(23):2197-206.
- Dankiewicz J, Cronberg T, Lilja G, Jakobsen JC, Levin H, Ullén S, et al. Hypothermia versus normothermia after out-of-hospital cardiac arrest. *New England Journal of Medicine* 2021;384(24):2283-94.
- Behringer W, Böttiger BW, Biasucci DG, Chalkias A, Connolly J, Dodd C, Khoury A, Laribi S, Leach R, Ristagno G. Temperature control after successful resuscitation from cardiac arrest in adults: A joint statement from the European Society for Emergency Medicine and the European Society of Anaesthesiology and Intensive Care. *Eur J Anaesthesiol*. 2024 Apr 1;41(4):278-281.
- Arrich J, Schütz N, Oppenauer J, Vendt J, Holzer M, Havel C, Herkner H. Hypothermia for neuroprotection in adults after cardiac arrest. *Cochrane Database Syst Rev*. 2023 May 22;5(5):CD004128.
- Granfeldt A, Holmberg MJ, Nolan JP, Soar J, Andersen LW; International Liaison Committee on Resuscitation ILCOR Advanced Life Support Task Force. Temperature control after adult cardiac arrest: An updated systematic review and meta-analysis. *Resuscitation*. 2023 Oct;191:109928.

-
14. Semeraro, Federico, Monsieurs, Koenraad G. et al. Synthesizing and communicating scientific evidence into treatment recommendations using artificial intelligence: The “case” of temperature control after cardiac arrest. *Resuscitation*, Volume 197, 110145.
 15. Andersen LW, Holmberg MJ, Nolan JP, Soar J, Granfeldt A. Conflicting guidelines: a commentary on the recent European Society for Emergency Medicine and European Society of Anaesthesiology and Intensive Care guidelines on temperature control after cardiac arrest. *Eur J Anaesthesiol*. 2024 Jul 1;41(7):468-472.
 16. Hjortrup PB, Haase N, Wetterslev J, Perner A. Gone fishing in a fluid trial. *Crit Care Resusc*. 2016 Mar;18(1):55-8. PMID: 26947417.
 17. Couper K, Andersen LW, Drennan IR, Grunau BE, Kudenchuk PJ, Lall R, Lavonas EJ, Perkins GD, Vallentin MF, Granfeldt A; International Liaison Committee on Resuscitation Advanced Life Support Task Force. Intraosseous and intravenous vascular access during adult cardiac arrest: A systematic review and meta-analysis. *Resuscitation*. 2025 Feb;207:110481.

Summary

Can we use artificial intelligence for creating and evaluating evidence based clinical guidelines in the field of cardiopulmonary resuscitation?

Artificial intelligence (AI) is without doubt on the rise in the field of medicine. Headlines such as “Doctors are incompetent and will be replaced by AI” is not unheard of. On the other hand, there are many caveats. AI can screen through massive amounts of data and make summaries. However, whether AI is able to make correct and relevant synthesis of the findings and capture the most important aspects of a specific medical question, is currently unclear. A recent example from the field of post-cardiac arrest care suggested fallacies in the reasoning of AI with regards to the evidence behind a particular treatment i.e. “therapeutic hypothermia”.