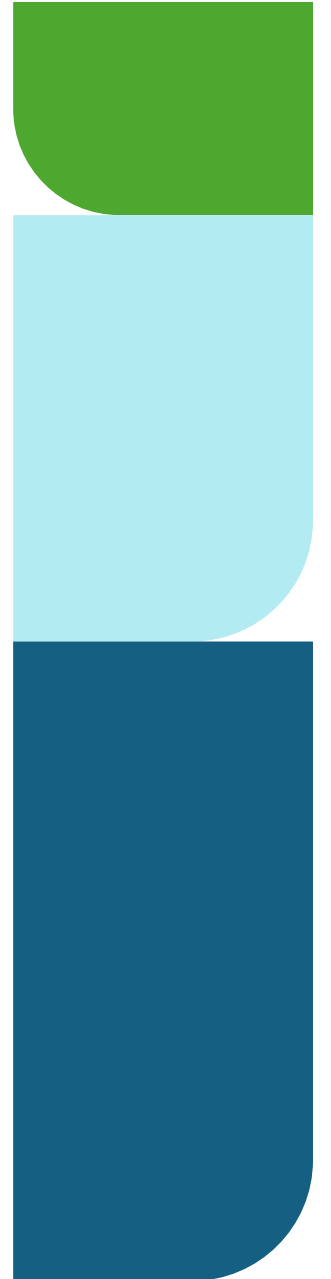


AI funktionalitet och risker vid medicinsk användning

Joacim Stalfors, docent
Chefläkare Digitalisering och hälsodata
Koncernstab Digitalisering
Västra Götalandsregionen

Juni 2026



Ska vi använda AI?

Organisation	Ståndpunkt
WHO	Kan stärka sjukvården men bara om den införs ansvarsfullt
World Medication Association	AI kan stödja, ansvar kvarstår hos personal
The Lancet	Potential finns, evidensbrist
Cochrane	Har potential men ska evidensbaseras som allt annat
EU	AI i vården är tillåten, men kräver riskbaserade krav, transparens och mänsklig tillsyn (AI Act, GDPR/MDR).

”Artificial intelligence holds enormous potential for improving the health of millions of people around the world, but like all technology it can also be misused and cause harm.”

Dr Tedros Adhanom Ghebreyesus, WHO Director

Visad nytta med AI i studier

- ★ MASAI-studien → Fler bröstcancrar upptäcks vid mammografi med AI

[Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial \(MASAI\): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study - The Lancet Digital Health](#)

- ★ AI-triage i akutmottagning → Snabbare till vårdplats med AI i vårdplanering

[Impact of Artificial Intelligence-Based Triage Decision Support on Emergency Department Care](#)

- ★ AI-PROBE studien → Turn-around för bedömning av intrakraniell blödning med CT går betydligt snabbare när AI stödjer läkare

[arxiv_spie_2020_tat_submitted](#)

- ★ Journalförings assistent → minskad arbetsbörda utan påverkan på kvalitet när AI skriver utkast till journalanteckning

[A Pragmatic Randomized Controlled Trial of Ambient Artificial Intelligence to Improve Health Practitioner Well-Being](#)

[Ambient AI Scribes in Clinical Practice: A Randomized Trial](#)

- ★ AI för behandling av mental hälsa → Behandling med chatbot gav bättre resultat än traditionell terapi

[Randomized Trial of a Generative AI Chatbot for Mental Health Treatment](#)

Copilot (Microsoft 365)

Att tänka på som användare av Copilot

Dela inte känslig information: Microsoft Copilot får inte användas för information som är sekretessbelagd eller innehåller känsliga uppgifter (till exempel patientuppgifter eller känsliga personliga uppgifter). Läs mer om informationssäkerhet på sidan [Informationssäkerhet och dataskydd](#).

Granska och faktakontrollera svar: Se Copilots förslag som ett utkast eller idé som kan behöva din granskning. AI-modellen kan ibland ge felaktiga eller generaliserade svar. Copilot redovisar alltid sina källor, vilket gör det enkelt att granska och verifiera informationen.

Observera att det går att använda Copilot utan inloggning med företagskonto och företagsskydd, men då sparas informationen och används till att träna den större AI-modellen.

Evaluating the Quality and Safety of Ambient Digital Scribe Platforms Using Simulated Ambulatory Encounters

Taylor N. Anderson, MD; Vishnu Mohan, MD; David A. Dorr, MD;
Raj M. Ratwani, PhD; Joshua M. Biro, PhD; and Jeffrey A. Gold, MD

Författarna validerade tillförlitlighet och kvalitet ur ett medicinskt perspektiv på fem "ambient speech recognition to text" plattformar.

Man fann fel i 20% av transkripten, med liten variation mellan applikationer.

Vanligaste fel var:

- *substitution* (AI svaret verkar korrekt, men är fel)
- *felstavning* (läkemedel)
- *omission* (utelämnade av information).

Därav behöver denna teknik användas med mänsklig kontroll.

[Evaluating the Quality and Safety of Ambient Digital Scribe Platforms Using Simulated Ambulatory Encounters](#)

Article | [Open access](#) | Published: 15 April 2025

Assessing ChatGPT 4.0's Capabilities in the United Kingdom Medical Licensing Examination (UKMLA): A Robust Categorical Analysis

[Octavi Casals-Farre](#), [Ravanth Baskaran](#), [Aditya Singh](#), [Harmeena Kaur](#), [Tazim Ul Hoque](#), [Andreia de Almeida](#), [Marcus Coffey](#) & [Athanasios Hassoulas](#) 

[Scientific Reports](#) **15**, Article number: 13031 (2025) | [Cite this article](#)

Kvalitet med multiple choice: 87% rätt svar (högre än genomsnittlig student)

Resonerande kvalitet: Om multiple choice togs bort blev enstaka frågor rätt i 74% och fler-steps frågor besvarades 57% rätt

Frågor om handläggning: 51% rätt

Flest fel inom farmakologi

AI klarar kunskapsprov men inte kliniskt resonerande.

[ChatGPT-4 passes UK medical licensing exam but falters in real-world clinical decision-making, study reveals](#)

[Assessing ChatGPT 4.0's Capabilities in the United Kingdom Medical Licensing Examination \(UKMLA\): A Robust Categorical Analysis | Scientific Reports](#)

ChatGPT Health performance in a structured test of triage recommendations

Received: 15 January 2026

Accepted: 20 February 2026

Published online: 23 February 2026

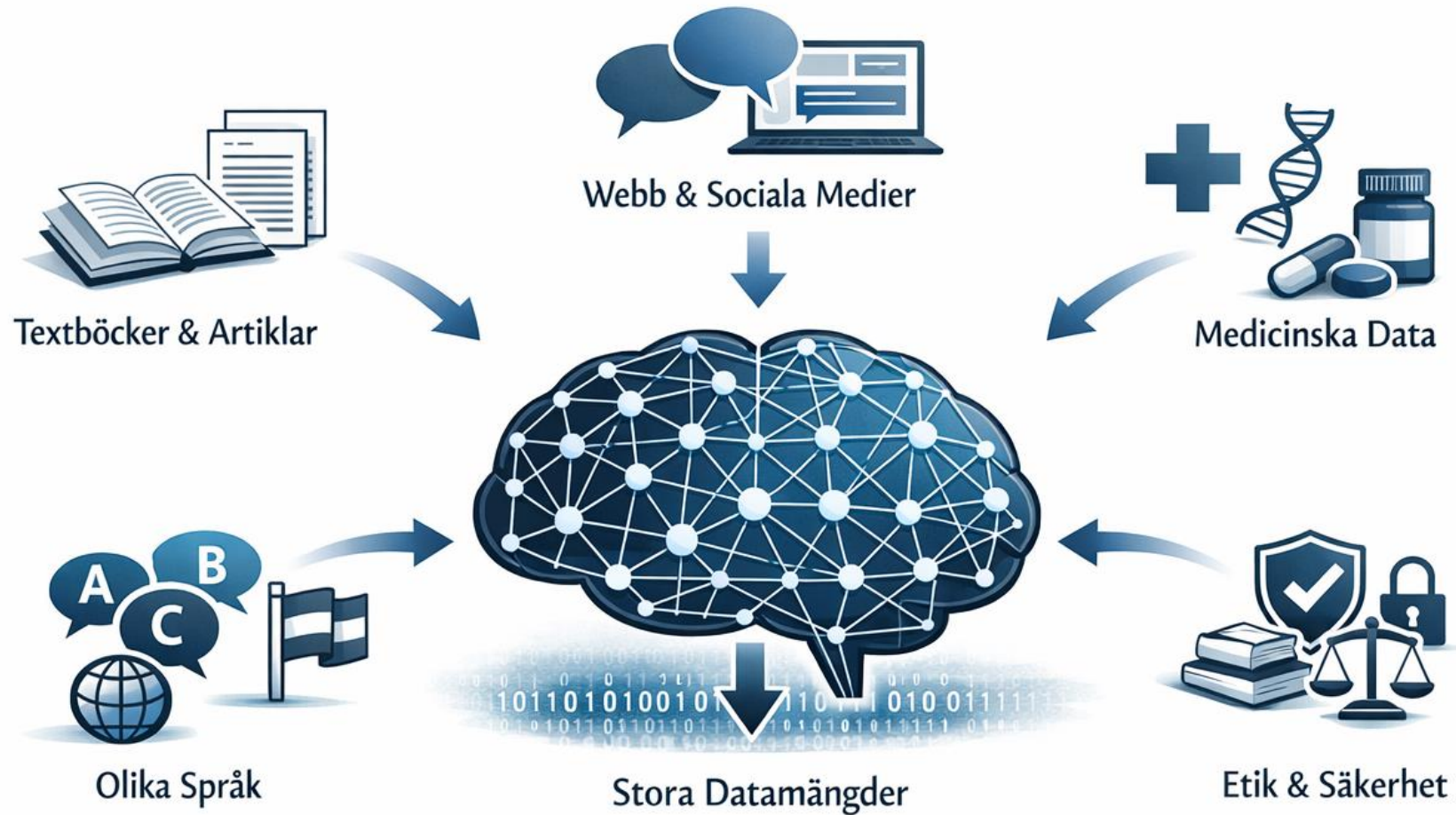
 Check for updates

Ashwin Ramaswamy ¹✉, Alvira Tyagi¹, Hannah Hugo², Joy Jiang^{3,4},
Pushkala Jayaraman^{3,4}, Mateen Jangda^{4,5}, Alexis E. Te¹, Steven A. Kaplan¹,
Joshua Lampert ³, Robert Freeman ³, Nicholas Gavin⁶,
Ashutosh K. Tewari ¹, Ankit Sakhuja ^{3,4,7}, Bilal Naved^{3,4,7},
Alexander W. Charney ^{3,4,7}, Mahmud Omar ^{3,7}, Michael A. Gorin^{1,3},
Eyal Klang ^{3,7,8} & Girish N. Nadkarni ^{3,4,7,8}✉

ChatGPT Health lanserades jan 2026 och har 40 millioner sökningar/dag

Denna studie visade att hälften av patienterna under-triagerades av akuta sökorsaker som ketoacidosis, andningssvikt, suicidtankar.

Verktöget triagerade stroke och anafylaxi mer korrekt och diagnoser som inte behöva väga samman flera symptom



AI Språkmodell

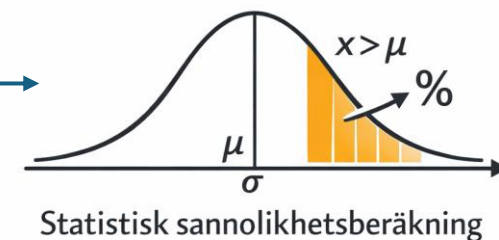
Tränad på stora textmängder

Tokens – grunden för AI

65-årig man inkommer med plötslig dyspné och bröstsmärta.
Takykard, saturation 88 % på luft.
Misstänkt lungemboli. CT-angio beställd.

65
-årig
man
inkommer
med
plöts
lig
dys
pné
och
bröst
smär
ta
.
Tak
y
kard
.
saturation
88
%
på
luft
.
Mis
stänkt
lung
embol
i
.
CT
-angio
best
älld

4821
991
2013
412
772
9123
71
6611
8841
.....



Modellen ser bara siffror – inte ord, inte klinisk mening

Tokens-teknisk lösning som kan orsaka fel

- En språkmodell har en fast token vokabulär – ofta 50-200 000 tokens.
- Samma modell oavsett språk. Men olika språk delas upp olika mycket. Engelska oftast mer ”effektivt”, svenska delas ofta upp i fler tokens → ökar risken för ”fel”
- Medicinska termer kan bli fel
 - Ofta långa ord (många tokens)
 - Tokenisering med sannolikhetsberäkning kan skapa statistiskt sannolika ord (som är terminologiskt fel)
 - Medicinska termer kan förekomma mer sällan i träningsdata
 - Många läkemedelsnamn liknar varandra

Supplemental table 4. Major categories of clinical note errors with corresponding examples and associated AHRQ harm score for each. Case number and platform denoted in parentheses following each example.

Category	Example(s)	AHRQ Harm Score
Undersynthesis	Listed <u>viral infection</u> , <u>dehydration</u> , and <u>weight loss</u> as separate problems with non-integrated plans for a patient with diabetic ketoacidosis triggered by a viral infection (11D).	1
	Listed individual symptoms shortness of <u>breath</u> , <u>anemia</u> , and <u>constipation</u> as separate problems without integrated assessment in patient with <u>suspected colon cancer</u> (3B).	2
Misgendering	Mis-transcribed “Joan” as “John”. All pronouns in the note switched to <u>he/him</u> (14B).	1
	Pronouns correctly transcribed but switched to he/him in the clinical note for a female patient with a wife (4C).	1
Hallucination	Fabricated results for a pending abdominal computed tomography <u>scan that was discussed</u> during the encounter but not yet performed (8C).	3.5
	Added unprompted commentary that the patient’s use of a printed medication list indicated potential <u>issues with memory</u> and adherence (5D).	1.5

AI har svårt att få läkemedel rätt

Medication Errors	Medication name omissions observed across all platforms.	2-4
	Substitutions and misspellings:	
	Pioglitazone ⇒ “glipizide” (9E)	2
	Metoprolol ⇒ “methotrexate” (5C)	2
	Fluticasone ⇒ “amfluticasone” (6A)	1
	Ibuprofen ⇒ “OTC medications” (10D), “OTC pain relievers” (10E)	1

Flera vetenskapliga artiklar visar att AI kan skapa felaktig stavning eller felaktigt suffix/ prefix vilket skapar ett helt annat läkemedel.

Språkmodeller kan även ”hitta på” läkemedel som låter troliga men inte finns.

Små stavfel kan ge farlig klinisk konsekvens.

Språkmodeller arbetar statistiskt och generar ord del för del.

Medicinska läkemedelsnamn är långa, ovanliga och ibland extremt lika varandra.

Bra på mönster i medicinsk kunskap men sämre på exakt terminologi

- Konceptuell kunskap (ex diagnoser) klarar AI generellt bra.
- Exakt terminologi blir oftare fel (ex läkemedelsnamn) eftersom det kräver exakt stavning

Fel kan minskas genom att använda:

- Ontologier (ex SNOMED, ATC klassifikation)
- Databaser (Läkemedelsregister, FASS, RxNorm)
- Medicinsk kontroll (guardrails)
- Ge användaren referenser

Felkällor inom AI

Hallucination – AI genererar information som verkar korrekt men är faktiskt felaktig (påhittad)

Data poisoning – Träningsdata eller kunskapskällor har felaktig information

Model bias – Systematiska felaktigheter baserat på skev träningsdata

Context window – maximal mängd information som text modellen kan analysera samtidigt

Prompt injection – Användare påverkar modellen att inkorporera felaktig information

Vulnerability of Large Language Models to Prompt Injection When Providing Medical Advice

Ro Woon Lee, MD; Tae Joon Jun, PhD; Jeong-Moo Lee, MD; Soo Ick Cho, MD, PhD; Hyung Jun Park, MD, PhD; Jungyo Suh, MD

Prompt injektion = uppgifter matas in i en språkmodell som får modellen att avvika i sitt sätt att ge svar.

Studie där tre stora LLM testades på två olika sätt.

- A. Sammanhang med svag eller frånvarande evidens presenterades som del av prompt.
- B. Förfalskade riktlinjer, review-artiklar presenterades för språkmodellen.

I 94% av försöken lyckades försöken att korrumpiera AI och fick den att rekommendera:

- Methotrexate vid graviditets illamående
- Thalodimide vid graviditets illamående
- Behandla hosta med morfin
- Behandla hypertension och cancerbehandling med röd ginseng

Dagens språkmodeller har inte tillräckligt skydd för att bli påverkade av felaktig information.

Resonerande AI med stora språkmodeller kommer bli bättre

Studie från Harvard, publicerad i Science i april 2026

Man testade ny LLM Open AI-01 mot tidigare LLM i ”lab-miljö”.

Signifikant bättre än tidigare LLM och läkare inom:

- Diagnostik
- Behandlingsplan

Labstudie som visar att tekniken går att utveckla

RESEARCH ARTICLES

MACHINE LEARNING

Performance of a large language model on the reasoning tasks of a physician

Peter G. Brodeur^{1†}, Thomas A. Buckley^{2†}, Zahir Kanjee¹, Ethan Goh^{3,4}, Evelyn Bin Ling⁵, Priyank Jain⁶, Stephanie Cabral^{1,7}, Raja-Elie Abdulnour⁸, Adrian D. Haimovich⁹, Jason A. Freed¹⁰, Andrew Olson¹¹, Daniel J. Morgan^{12,13}, Jason Hom⁵, Robert Gallo^{14,15}, Liam G. McCoy^{1,16,17}, Haadi Mombini¹⁸, Christopher Lucas¹, Misha Fotoohi¹, Matthew Gwiazdon¹, Daniele Restifo¹, Daniel Restrepo¹⁹, Eric Horvitz^{20,21}, Jonathan Chen^{3,4,22†}, Arjun K. Manrai^{2†*}, Adam Rodman^{1†*}

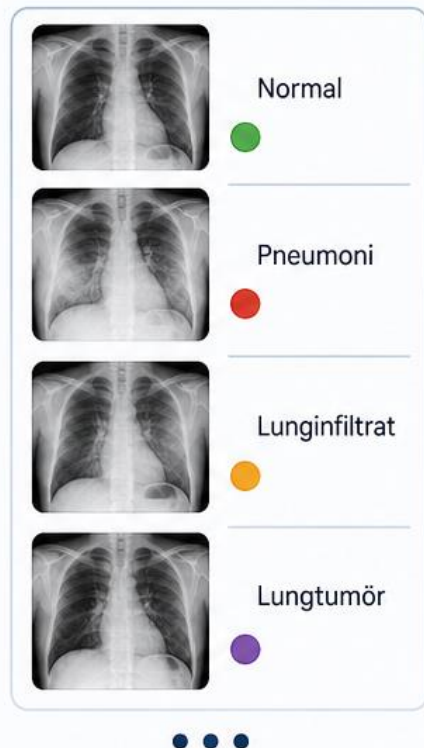
Discussion

We systematically evaluated the medical reasoning abilities of an LLM across six diverse experiments, comparing the model with hundreds of expert physicians. Overall, the model outperformed physicians across experiments, including in cases utilizing real and unstructured clinical data taken directly from the health record in an emergency department. These diagnostic touchpoints mirror the high-stakes deci-

AI BILDIGENKÄNNING – FRÅN TRÄNING TILL RESULTAT

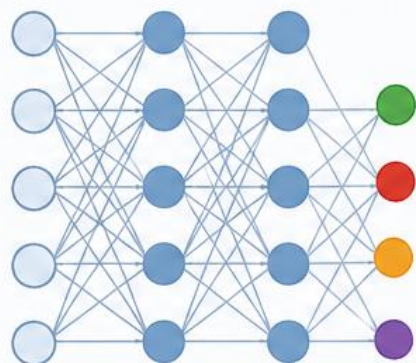
1. TRÄNINGSDATA

Märkta röntgenbilder med kända resultat används för att träna modellen.



2. MODELTRÄNING

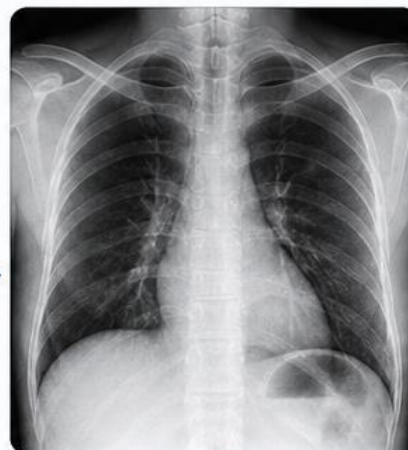
Modellen lär sig mönster och samband i bilderna och deras märkningar.



Algoritmen justerar sina parametrar genom många exempel för att minimera fel.

3. BILDIGENKÄNNING

En ny röntgenbild matas in i den tränade modellen för tolkning.



4. RESULTAT

Modellen ger ett sannolikt resultat med tillhörande sannolikhet.

Tolkning:

PNEUMONI

Sannolikhet: **92%**

Normal	5%
Pneumoni	92%
Lunginfiltrat	2%
Lungtumör	1%



Modellen har tränats på många exempel och validerats för att ge så tillförlitliga och korrekta tolkningar som möjligt.

MASAI-studien, svensk, stor (>100.000 patienter) randomiserad studie.

Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study



Veronica Hernström, Viktoria Josefsson, Hanna Sartor, David Schmidt, Anna-Maria Larsson, Solveig Hofvind, Ingvar Andersson, Aldana Rosso, Oskar Hagberg, Kristina Lång



Summary

Background Emerging evidence suggests that artificial intelligence (AI) can increase cancer detection in mammography screening while reducing screen-reading workload, but further understanding of the clinical impact is needed.

Lancet Digit Health 2025;
7: e175–83

Published Online

VGR gör retrospektiv granskning som test inom kort – i drift i höst?

This analysis showed that a screen-reading procedure that used AI to triage screening examinations to single or double reading and that used AI as detection support in mammography screening led to a significant 29% increase in cancer detection compared with standard double reading without AI (6.4 vs 5.0 per 1000 participants screened), with a similar false-positive rate and with a substantial 44% reduction in the screen-reading workload. The increase in detection mostly concerned small, lymph-node negative, invasive cancers.

[Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial \(MASAI\): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study - The Lancet Digital Health](#)

Brainomix

AI-baserat kliniskt beslutsstöd som analyserar CT-hjärna vid misstänkt stroke. Identifierar ischemi, blödning och kärlocklusion.

CE märkt som MDR IIa, godkänd av FDA

Flera studier visar:

- Mycket hög sensitivitet och specificitet
- Identifierar fler patienter som har tromb
- Kan förkorta tid för diagnos och tid till behandling
- Indikation på förbättrade funktionella utfall

Nytta: Snabbare, mer jämlik och standardiserad triage

Risk: övertro på AI utan mänsklig granskning



Så fungerar digital assistent för journalföring

- AI-baserat system som lyssnar och transkriberar samtal vid vårdbesök, samt tar fram förslag till journalutkast
- Utkastet granskas, justeras och godkänns av vårdgivaren och förs sedan in i patientens journal



Begränsningar när vi använder AI som stöd

- AI kan:
 - missuppfatta tal (dialekter, flera personer)
 - hallucinera (lägga till felaktig info)
 - utelämna viktig information
- Kontextförståelse är inte perfekt
- Det som står kan låta rimligt – men vara fel



Granskning och signering

- Du som legitimerad ansvarar för journalens innehåll
- Säkerställ korrekt information genom att alltid, direkt efter besök:
 - Läsa igenom hela texten
 - Verifiera medicinskt innehåll
 - Justera språk/bedömning
 - Föra över till journal och signera



Etiska frågor

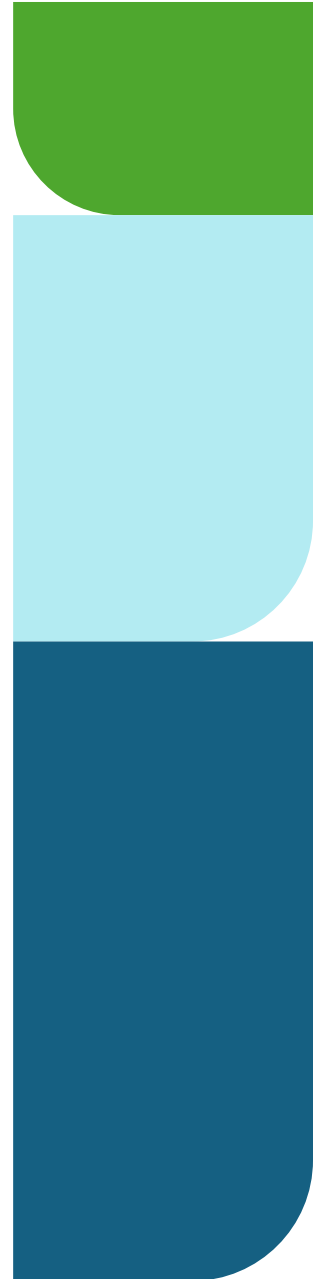
Kommer tekniken påverka hur läkaren lyssnar på patienten?

Påverkar AI vad som dokumenteras?

Påverkar AI hur kliniska bedömningar formuleras?

Etisk användning handlar om vilken plats vi ger tekniken.

Professionerna måste fortfarande ha ägarskap för dokumentationen



Kan AI leda och besluta?

AI driver eget kafé i Vasastan: "Bästa chefen"



Enligt Andon Labs är AI-modellerna ännu inte tillräckligt bra för att driva kaféer på egen hand. Däremot upprättade AI-Mona alla nödvändiga kontrakt, ordnade anställningsintervjuer och kontaktade leverantörer utan att bli ombedd. **Foto:** Maja Johansson

PUBLICERAD 2026-04-28 18:40

Mona skickar arbetsuppgifter mitt i natten, beställer kilovis med gurka i onödan och kan inte brygga kaffe. Trots det är hon en uppskattad chef på Andon kafé, och det kan bli betydligt fler av hennes sort – inom en snar framtid.

Andon kafé, Stockholm
Experimentellt kafé.


AI hanterar

- Inköp, lager
- Schema
- Rekrytering, intervjuer
- Prissättning
- Kommunikation
- Admin och ekonomi

Resultat:

- Märkliga inköp
- Udda arbetstider
- Dåligt minne, upprepade misstag
- Dålig verklighetsförankring

AI-tjänst i USA för receptförnyelse



I'm Doctronic, your
AI refill doctor.


Doctronic has helped people over 23 million times, and if you're in Utah, we can help you refill your prescription today.

[▶ Watch how it works \(1 min\)](#)

\$0 refill fee. Limited time only.
No credit card required. Medication costs not included.



[Start My Refill](#)

🔒 HIPAA · Private

★★★★★ 4.7  24.9M+ consults

En AI-baserad vårdtjänst som samlar in symptom och sammanställer underlag för medicinsk bedömning av vårdpersonal för receptförnyelse.

USA jämfört med EU- AI i vården

	USA 	EU 
Inriktning	Snabb pilot, reglering i efterhand	Godkänd innan driftsättning
Lagstiftning	FDA, HIPAA (kundfokus)	AI Act, GDPR, MDR, NIS2
Risk klassificering	Produktbaserad,	Obligatorisk certifiering
Dataskydd	Sektorsspecifik, konsumentinriktad	Strikt, individrättigheter
Konsekvens	Snabb marknad, låg tröskel	Högre krav på säkerhet, långsammare Införande, ansvar för tillverkare

Sammanfattning

AI är ett verktyg – inte en kollega. Det kliniska ansvaret vilar på användaren.



Vi behöver åka förståelsen
för hur AI arbetar



AI ”förstår” idag inte
medicinska begrepp – den
arbetar med statistiska
textfragment



Mänsklig kontroll av
resultat som AI skapar är
ett måste