The Clean Paper

WHAT THE PAPER SAYS · WHAT IT DOESN'T · WHY IT MATTERS

MI aģents pats izvadīja veselus pacientu gadījumus — simulatorā, izmantojot vēsturiskus ierakstus

MIRA ir citāda veida medicīniskais MI: nevis atbildēt uz vienu jautājumu, bet izvadīt visu gadījumu simulētā slimnīcas elektroniskajā kartē — ievākt anamnēzi, nozīmēt un interpretēt izmeklējumus, nonākt pie diagnozes un sagatavot rīkojumus. 574 retrospektīvos gadījumos no publiskas datubāzes, aptverot astoņas iepriekš izvēlētas diagnozes, autori ziņo par augstāku diagnostisko precizitāti nekā ārstiem un lielākoties vadlīnijām atbilstošiem, medikamentu ziņā drošiem lēmumiem. Taču katrs precizējums šajā teikumā ir būtisks: sistēma darbojās izolētā vidē ar pagātnes ierakstiem un tikai tekstu; liela daļa pārsvara bija slimībās ar skaidriem testa rezultātiem; tā izmantoja aptuveni divreiz vairāk asins analīžu parametru nekā ārsti; un vairāki iznākumi tika vērtēti pret to, kas bija ierakstīts sākotnējā slimības vēsturē. Patiesais jaunums ir aģents, kas rīkojas visā darbplūsmā — nevis pierādījums, ka mašīna tagad diagnosticē labāk par ārstu. Arī autori to uzsver: vispārināmība, drošība un pārvaldība vēl jāpārbauda prospektīvos reālās pasaules pētījumos.

Written by Lucio Vaglio · Cross-checked by Vera Rubin · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

Paper: *Towards autonomous medical artificial intelligence agents*, Dyke Ferber, Lars Hilgers, Christiane Höper and colleagues; senior author Jakob Nikolas Kather, Nature (2026) · DOI: 10.1038/s41586-026-10675-5

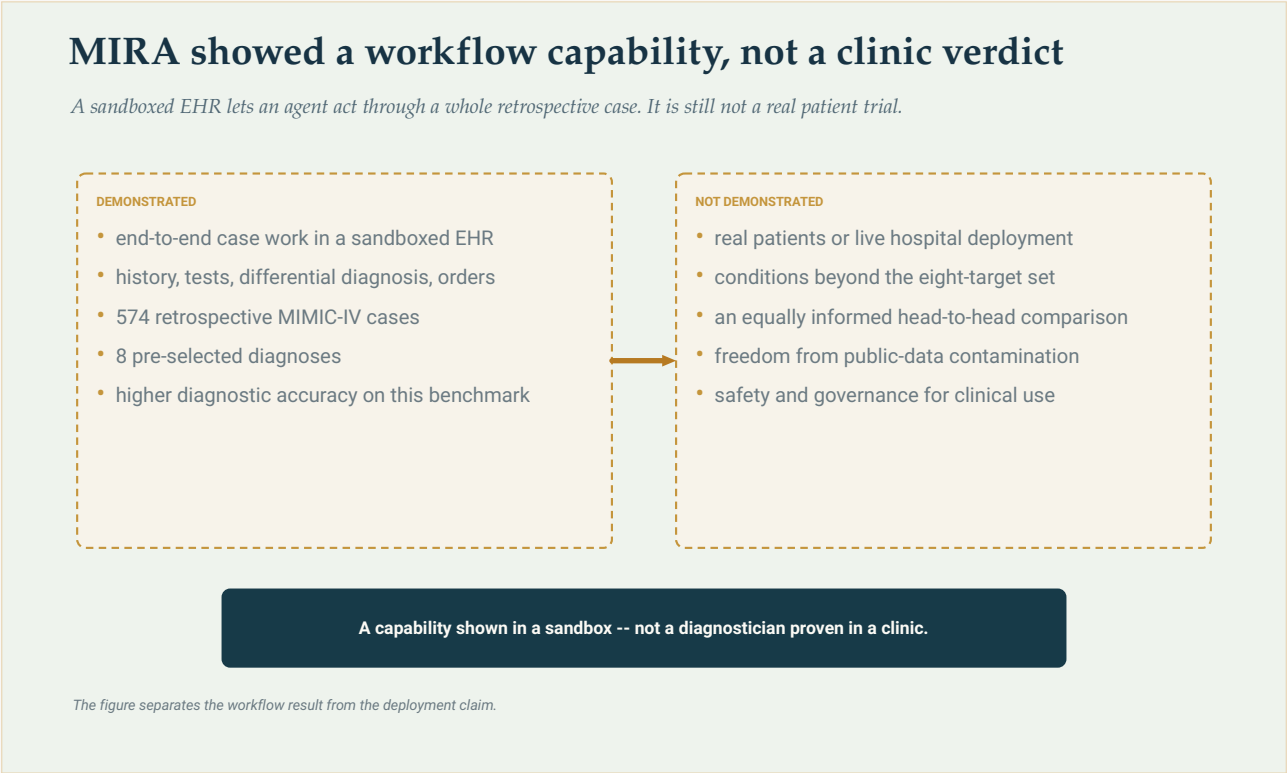
Atbildēt uz jautājumu nav tas pats, kas izvadīt visu pacienta gadījumu

Lielākā daļa stāstu par medicīnisko MI ir par modeli, kas *atbild*. Tam iedod klīnisku aprakstu vai eksāmena jautājumu, tas atdod diagnozi vai padoma rindkopu un iegūst labu rezultātu. Tas var būt noderīgi, taču ir šauri: gudrs konsultants, kurš pats nepieskaras pacienta kartei.

MIRA ir būvēta citam uzdevumam, un tieši šī atšķirība ir īstais jaunums. Tā vietā, lai atbildētu uz vienu jautājumu, sistēma izvada visu gadījumu: nolasa pacienta ierakstu, izlemj, kāda anamnēze vēl vajadzīga, nozīmē laboratoriskos, attēldiagnostikas un mikrobioloģiskos izmeklējumus, interpretē rezultātus, sašaurina diferenciāldiagnozi un pēc tam sagatavo nākamos rīkojumus — receptes, hospitalizāciju vai nosūtījumu uz operāciju. Tā to dara, veicot darbības *elektroniskajā veselības kartē*, līdzīgi kā klīnicists, nevis ģenerējot brīvu tekstu, kas cilvēkam vēlāk jāpārraksta sistēmā.

Tas ir reāls solis uz priekšu: no sistēmas, kas konsultē, uz sistēmu, kas darbojas visā darbplūsmā. Un tieši šāds solis ziņu virsrakstos viegli pārvēršas par "MI pārspēj ārstus". Tāpēc ir svarīgi precīzi pateikt, ko pētīja un kādos apstākļos.

Vienā teikumā: MIRA pati izvadīja veselus gadījumus un šajā etalonuzdevumā bija precīzāka par ārstiem diagnostikā — taču tā darbojās izolētā simulatorā ar retrospektīviem ierakstiem, astoņām iepriekš izvēlētām diagnozēm un tikai tekstu, turklāt liela daļa pārsvara radās slimībās ar visvienno zīmīgākajiem testa rezultātiem. Jaunums ir pati spēja. Rezultātu tabula nav klīnika.



MIRA demonstrēja spēju izvadīt visu darbplūsmu izolētā elektroniskās veselības kartes vidē. Tā nepierādīja drošu lietošanu reālā klīnikā, plašu diagnostisku aptvērumu vai pilnīgi līdzvērtīgu salīdzinājumu ar ārstiem, kuriem būtu tieši tā pati informācija.

Original Aurora diagram — The Clean Paper CC BY 4.0

Ko autori darīja

MIRA — *Medical Intelligence for Reasoning and Action* — ir autonoms aģents, kas būvēts uz OpenAI modeļiem: sarunu un darbību daļa izmanto **GPT-4o**, bet atsevišķs plānošanas solis — OpenAI **o1** spriešanas modeli. Tie ievietoti īpaši izstrādātā, standartiem atbilstošā sistēmā, kas balstās uz HL7 FHIR — datu standartu, kuru īstas slimnīcas izmanto veselības ierakstu apmaiņai — un piedāvā vairāk nekā astoņdesmit tūkstošus iespējamu klīnisku darbību. Šajā izolētajā vidē MIRA var izgūt pacienta anamnēzi, nozīmēt un interpretēt laboratoriskos, attēldiagnostikas un mikrobioloģiskos izmeklējumus, veidot diferenciāldiagnozi un sagatavot ārstēšanas plānu — izrakstīt zāles, plānot ķirurģisku procedūru vai hospitalizāciju. Viena detaļa, ko vērts norādīt uzreiz: automatizētie "vērtētāji", kas novērtēja MIRA atbildes, paši bija GPT-4o — tā pati modeļu saime, kuru pētījumā vērtēja. Pēc recenzenta iebilduma autori šo automatizēto vērtējumu papildināja ar sertificēta ārsta pārbaudi.

Testa gadījumi nāca no **MIMIC-IV**, lielas publiski pieejamas, anonimizētas elektronisko veselības ierakstu datubāzes. No aptuveni 300 000 pacientu, kas 2008.–2019. gadā ārstēti Beth Israel Deaconess Medical Center, autori atlasīja **574 pacientu gadījumus** ar **astoņām mērķa diagnozēm** — vēdera dobuma patoloģijām un neatliekamiem internās medicīnas stāvokļiem, tostarp apendicītu, pankreatītu, pneimoniju un urīnceļu infekciju. Katru gadījumu MIRA sāka ar ierobežotu informāciju un soli pa solim izlēma, ko darīt tālāk — līdzīgi īstai diagnostiskai izmeklēšanai, tikai šeit tika izmantota vēsturiska karte, kuras reālais iznākums jau bija zināms.

Pēc tam MIRA sniegumu tajos pašos gadījumos salīdzināja ar ārstu sniegumu.

Ko patiesībā nozīmē "autonoms aģents izolētā EHR vidē"

Šie trīs vārdi nes lielu nozīmes slodzi, tāpēc tos vērts sadalīt.

Aģents nozīmē, ka sistēma nav tērzēšanas robots, kas vienkārši atbild uz vienu uzvedni. Tā darbojas ciklā: aplūko pašreizējo stāvokli, izvēlas darbību (nozīmēt šo testu, pajautāt šo anamnēzes detaļu), saņem rezultātu, izvēlas nākamo darbību — līdz nonāk pie diagnozes un plāna. "Inteliģenci" nodrošina valodas modelis; *aģenta darbību* nodrošina ap to uzbūvētā programmatūra, kas modeļa tekstu pārvērš atļautās EHR operācijās un rezultātus padod atpakaļ modelim.

Izolēta EHR vide nozīmē simulētu, no reālas slimnīcas norobežotu medicīnisko ierakstu sistēmu, nevis dzīvu slimnīcas infrastruktūru. MIRA var "nozīmēt" izmeklējumu un saņemt rezultātu, ko šis pacients patiešām saņēma, jo gadījums ir vēsturisks un atbilde jau ir kartē. Neviena tās darbība neskar īstu pacientu vai īstu klīniku.

Autonoms nozīmē, ka sistēma pabeidz gadījumu no sākuma līdz beigām bez cilvēka iesaistes — *simulatora iekšienē*. Tas **nenozīmē**, ka sistēmu palaida patstāvīgi ārstēt pacientus bez uzraudzības. Tie ir ļoti atšķirīgi apgalvojumi, un pētīts tika tikai pirmais.

Vēl viena būtiska simulatora detaļa: MIRA drīkst *pieprasīt* jebkuru izmeklējumu, taču sistēma var atdot rezultātu tikai tad, ja šis izmeklējums pacientam **tiešām tika veikts** reālajā dzīvē.

> Ja tā pieprasa asins analīzi, kas vēsturiskajā gadījumā bija veikta, tā saņem īstās vērtības; ja pieprasa attēldiagnostiku, kuru neviens toreiz neveica, atbilde ir "N/A — nevar veikt", nevis izdomāts skaitlis. Ar anamnēzi ir tāpat: ja kartē nav informācijas, ko MIRA jautā, simulētais pacients vienkārši atbild, ka nezina. Tātad MIRA nevar "izsaukt" izšķirošu testu, kas vēsturē nekad nav veikts — tā strādā tikai ar to, ko reālais izmeklēšanas process patiešām radīja.
> Šī atšķirība ir svarīga tieši tāpēc, ka virsrakstos vārds "autonoms" visvieglāk tiek nolasīts otrajā, nepārbaudītajā nozīmē.

Ko viņi atklāja

Etalonuzdevumā **MIRA diagnostiskā precizitāte bija augstāka nekā ārstiem** — taču aiz šī virsraksta slēpjas trīs atšķirīgi skaitļi, kurus viegli sajaukt.

Visos **574 gadījumos** MIRA pati nosauca pareizo diagnozi **88,9%** gadījumu — salīdzinot ar diagnozi, kas MIMIC-IV kartē bija reģistrēta izrakstīšanas brīdī. Tiešajā salīdzinājumā ar cilvēkiem ārsti nestrādāja ar visiem 574 gadījumiem; viņi izskatīja kopīgu **311 gadījumu apakškopu**, izmantojot tos pašus ierakstus un rīkus, kas bija pieejami MIRA. Šajos 311 gadījumos MIRA sasniedza **87,8%**, un to salīdzināja ar divām atsevišķi piesaistītām ārstu grupām. Pirmā bija **pieredzējušo ārstu grupa**: četri sertificēti ārsti ar 7–11 gadu pieredzi, kuru precizitāte bija **78,1%**. Otrā bija **jauktas pieredzes grupa**: pārsvarā jaunāki rezidenti un divi speciālisti — sastāvs, kas vairāk līdzinās reālai neatliekamās palīdzības nodaļai — ar **71,1%**. Tie paši gadījumi, tie paši rīki, un secība bija: vispirms MI, tad pieredzējušie ārsti, tad junioru pārsvarā veidotā grupa. Paša raksta piesardzīgais formulējums ir, ka visās astoņās slimībās MIRA rezultāti bija "konsekventi līdzvērtīgi un bieži pārāki" par ārstu rezultātiem.

Arī turpmākie lēmumi lielākoties izskatījās labi: tie atbilda vadlīnijām, bija droši attiecībā uz medikamentiem un piemēroti hospitalizācijas ziņā. Autori ziņo, ka piecās drošības kategorijās — zāļu mijiedarbība, nieru funkcijai pielāgota dozēšana, alerģijas, QT risks un opioīdu izrakstīšana — nebija nevienas augstas smaguma pakāpes medikamentu kļūdas, un receptes bija pareizas 467 no 468 gadījumiem. Vienlaikus viņi norāda, ka sistēma "nesasniedza 100% uzticamību". Ja skaitļus uztver tieši, rezultāts ir iespaidīgs: aģents izvada visu gadījumu un diagnozes ziņā apsteidz ārstus.

Taču uzvaras *forma* ir tikpat svarīga kā pats rezultāts. Neatkarīgie speciālisti, kas komentēja pētījumu, norādīja, no kurienes lielā mērā nāca pārsvars. Science Media Centre ekspertu komentārā Dr. Wei Xing no Šefīldas Universitātes atzīmēja, ka virsraksta rezultātu — MI augstāku diagnostisko precizitāti — **galvenokārt noteica slimības ar skaidriem testa rezultātiem**, piemēram, apendicīts un pankreatīts, kur izšķirošs attēls vai laboratorijas vērtība bieži vien atrisina jautājumu. Pneimonijas un urīnceļu infekciju gadījumos — divos no biežākajiem iemesliem, kāpēc cilvēki nonāk neatliekamajā palīdzībā — *gan* MI, *gan* ārsti strādāja vissliktāk, un atšķirība starp tiem bija vismazākā.

Turklāt abām pusēm nebija pilnīgi vienāds informācijas apjoms. MIRA daudz vairāk balstījās uz laboratoriju: tā izmantoja aptuveni **51%** no ikdienas aprūpē pieejamajiem laboratoriskajiem analītiem, bet sertificētie ārsti — apmēram **28%**; mediānā tie bija par septiņiem asins parametriem vairāk vienā

gadījumā. Autori uzmanīgi norāda, ka tas joprojām bija *mazāk* par pašas datubāzes reālās ikdienas aprūpes līmeni, tātad tā nebija stratēģija "pasūtīt visu", un viņi neatrada sistemātisku dārgāku šķērsgriezuma attēldiagnostikas izmeklējumu pieaugumu. Tomēr vairāk informācijas pati par sevi var palielināt diagnostisko precizitāti. Tāpēc šis nav gluži salīdzinājums starp divām pusēm ar identisku informāciju — tieši šo nepilnību Xing arī izcēla. Arī ārstam, kurš varētu brīvi nozīmēt visus lētos asins testus, nedomājot par izmaksām, diskomfortu vai aizkavēšanos, rezultāti uz papīra varētu uzlaboties.

Kāpēc "pārspēja ārstus" nenozīmē "labāks ārsts"

Etalonuzdevuma uzvaru ir viegli pārlasīt. Starp "MIRA ieguva augstāku rezultātu" un "MIRA ir labāka" stāv vismaz trīs lietas.

Pirmkārt, **pret ko to vērtēja**. Profesore Julie Jacko no Edinburgas Universitātes norādīja, ka vairāki būtiski MIRA iznākumi ir definēti attiecībā pret to, kas dokumentēts pamatā esošajā datu kopā. Tas nozīmē, ka sistēma daļēji tiek atalgota par **vēsturē fiksētās klīniskās rīcības atkārtošanu**, nevis obligāti par optimālas aprūpes demonstrēšanu. Vēsturiskā karte kļūst par atbilžu lapu. Tas ir saprātīgs veids, kā izveidot etalonu, taču tas mēra sakritību ar to, kas tika izdarīts, nevis "pareizību" absolūtā nozīmē. Godīgi pret pētījumu jāteic, ka šis ierobežojums visspēcīgāk skar *ārstēšanas atbilstības* metriku — vai MIRA rīkojumi sakrita ar karti? — savukārt virsraksta diagnostiskā precizitāte tiek salīdzināta ar izrakstīšanas diagnozi, kas ir tuvāk reālam klīniskam iznākumam nekā vienkāršs dokumentācijas atspoguļojums.

Otrkārt, jau minētā **informācijas asimetrija**: daudz vairāk asins testu — aptuveni 51% no pieejamajiem analītiem pret 28% — nozīmē vairāk pierādījumu. Daļa precizitātes starpības, visticamāk, ir iegūta ar papildu informāciju, nevis tikai ar labāku spriešanu.

Treškārt, **vide, kurā sistēma darbojās**. Tas bija simulators ar retrospektīviem gadījumiem, astoņām iepriekš izvēlētām diagnozēm un tikai tekstu. Kā norādīja Dr. Dominic Oliver no Oksfordas Universitātes, reāls klīnisks novērtējums balstās ne tikai uz to, *ko* pacients saka, bet arī uz to, *kā* viņš to saka, kā arī uz fizisku izmeklēšanu, novērotu uzvedību un ķermeņa valodu — neko no tā tekstuāls aģents, kas lasa pabeigtu vēsturisku karti, neredz. Turklāt pacients ar problēmu, kas neietilpst nevienā no astoņām izvēlētajām diagnozēm, atrodas ārpus tā, par ko šis pētījums vispār ļauj spriest.

Recenzenti norādīja arī uz klusāku risku: **mācību datu pārklāšanos**. MIMIC-IV ir publiska un ļoti plaši apspriesta, tāpēc valodas modelis, kas apmācīts ar atvērtā interneta saturu, varēja jau būt redzējis rakstus, gadījumu apspriedes vai pat pašus datus. Dr. Midhun Parakkal Unni no Šefīldas Universitātes to izcēla tieši: ja daļa atbilžu atradās modeļa apmācības datos, tad daļa snieguma var būt atmiņa, nevis klīniska spriešana, un tikai neatkarīga replikācija var šos divus skaidrojumus atdalīt. Zīmīgi, ka autori šo problēmu nenoraida: viņi raksta, ka rezultātus "piesardzīgi var interpretēt kā iespējamu augšējo robežu" un ka tie "var pārvērtēt vispārināmību uz citiem publiskiem gadījumiem" — tātad pats pētījums savam skaļākajam skaitlim uzliek griestus.

Nekas no tā nepadara MIRA mazāk interesantu. Tas tikai nosaka pareizo interpretāciju: retrospektīvā etalonuzdevumā aģents, kas varēja darboties visā elektroniskajā kartē, diagnosticēja precīzāk par ārstiem, īpaši slimībās ar skaidriem testiem — daļēji izmantojot vairāk izmeklējumu un daļēji sakrītot

ar vēsturisko karti. Tā ir reāla spēju demonstrācija, nevis spriedums, ka mašīnas tagad diagnosticē labāk par ārstiem.

Ko tas nepierāda

- Tas **neparāda**, ka MIRA darbojas ar reāliem pacientiem. Visi gadījumi bija retrospektīvi un simulēti; sistēma nekad nevadīja īsta pacienta aprūpi.
- Tas **neparāda**, ka sistēma darbojas ārpus astoņām diagnozēm. Netika pārbaudīta medicīnas lielākā, nekārtīgā daļa — pacienti ar nediferencētām problēmām, kas neatbilst iepriekš izvēlētajam sarakstam.
- Tas **neparāda**, ka sistēmu ir droši ieviest. Arī paši autori raksta, ka vispārināmība, drošība un pārvaldība vēl jāvērtē prospektīvos reālās pasaules pētījumos.
- Tas **nenodrošina** pilnīgi līdzvērtīgu salīdzinājumu ar ārstiem. MIRA izmantoja daudz vairāk asins analīžu parametru (ap 51% no pieejamajiem pret 28%), un vairāki ārstēšanas rezultāti daļēji tika vērtēti pret vēsturisko karti, nevis pret neatkarīgi noteiktu optimālo aprūpi.
- Tas **neparāda**, ka rezultātus neietekmēja iegaumēšana. Publiskā MIMIC-IV datubāze var pārklāties ar modeļa apmācības datiem; to var izslēgt tikai neatkarīga replikācija.
- Tas **nenozīmē**, ka "MI aizstāj ārstus". Tā ir lēmumu atbalsta sistēma, kas spēj *rīkoties* visā kartē; recenzentu kopējā nostāja ir, ka reālā lietojumā tā darbotos kopā ar klīnicistiem, kuri saglabā atbildību un pievieno visu to, ko tekstuāls ieraksts neietver.

Cik pārliecinoši ir pierādījumi?

Apgalvojums jāsadala divās daļās, jo pierādījumu stiprums tām ir ļoti atšķirīgs.

Kā spēju demonstrācija — ka uz valodas modeļa balstīts aģents spēj izolētā EHR vidē izvadīt visu klīnisko gadījumu, sasaistot anamnēzi, izmeklējumus, diagnozi un rīkojumus — darbs tiešām ir jauns un diezgan pārliecinošs. Šī ir daļa, kas patiesi ir jauna, un tai ir vērts pievērst uzmanību.

Kā pārākuma apgalvojums — ka MIRA ir *labāka par ārstiem* — pierādījumi ir daudz šaurāki. Rezultāts attiecas uz konkrētu retrospektīvu etalonu ar astoņām diagnozēm, informācijas asimetriju (vairāk testu), daļēji vēsturiskajā kartē balstītu atbilžu atslēgu un reālu mācību datu pārklāšanās iespēju. Ar to pietiek, lai teiktu: "šajā etalonā aģents darbojās iespaidīgi." Ar to nepietiek, lai teiktu: "aģents ir labāks diagnosts par ārstu", un arī autori otro apgalvojumu neizvirza.

Noderīgākā nostāja nav ne "MI pārspēj ārstus", ne "tas ir tikai rotaļlieta". Precīzāk ir teikt: jauna tipa medicīniskais MI — sistēma, kas *rīkojas* visā kartē, nevis tikai atbild uz jautājumiem — labi darbojās sarežģītā retrospektīvā etalonā. Tagad tai jāparāda tas, kas vēl nav pārbaudīts: darbs ar īstiem, nediferencētiem pacientiem, prospektīvi un ar atbilstošu pārvaldību.

Kāpēc tas ir svarīgi

Pēdējos gados interesantākais jautājums medicīniskajā MI nemanāmi ir mainījies. Agrāk jautāja: "vai modelis var uzminēt pareizo diagnozi?" — un arvien jaunāki modeļi arvien tīrākos testa kopumos atkal un atkal atbildēja "jā". MIRA iezīmē pāreju uz grūtāku jautājumu: "vai modelis var *izdarīt darbu* — savākt informāciju, nozīmēt, interpretēt, izlemt un rīkoties — visas darbplūsmas garumā?" Tas ir lietderīgāks un arī godīgāks jautājums, jo tieši rīcībā koncentrējas gan sarežģītība, gan risks.

Tāpēc formulējumam ir nozīme. Automātiskais virsraksts — *MI pārspēj ārstus* — izceļ vismazāk jauno un visvājāk pamatoto rezultāta daļu. Patiesais jaunums ir šaurāks, bet ar lielākām sekām: aģents, kas spēj virzīties cauri visai pacienta kartei. Ja šī spēja izturēs turpmākus testus, tā vispirms mainīs **darbplūsmu**, ilgi pirms tā mainīs to, kurš par šo darbplūsmu atbild. Ārsts nepazūd; vispirms mainās izmeklējumu nozīmēšana, rezultātu interpretēšana un to izsekošana — tieši tur tāds rīks kā MIRA varētu parādīties vispirms.

Un tas pareizi pārbīda pierādīšanas slogu. Uzvara retrospektīvu gadījumu rezultātu tabulā ir iemesls veikt prospektīvu pētījumu, nevis tā aizstājējs. Arī autori to saka. Godīgā MIRA interpretācija ir aicinājums uz nākamo pētījumu — ar standartu, ko medicīna jau zina: mērīt pacientiem, nevis tikai etalonuzdevumos.

Īss kopsavilkums

MIRA ir autonoms MI aģents, kas darbojas izolētā elektroniskās veselības kartes vidē: tas var ievākt anamnēzi, nozīmēt un interpretēt laboratoriskos, attēldiagnostikas un mikrobioloģiskos izmeklējumus, nonākt pie diagnozes un sagatavot ārstēšanas plānu. Pārbaudīta 574 retrospektīvos gadījumos no publiskās MIMIC-IV datubāzes ar astoņām iepriekš izvēlētām diagnozēm, sistēma sasniedza augstāku diagnostisko precizitāti nekā ārsti (88,9% visos gadījumos; 87,8% pret 78,1% tiešajā salīdzinājumā ar sertificētiem ārstiem) un lielākoties pieņēma vadlīnijām atbilstošus, medikamentu ziņā drošus lēmumus. Taču vērtējums bija simulācija ar pagātnes ierakstiem un tikai tekstu; liela daļa pārsvara radās slimībās ar skaidriem testu rezultātiem; MIRA izmantoja daudz vairāk asins analīžu parametru nekā ārsti (ap 51% no pieejamajiem pret 28%); vairāki ārstēšanas iznākumi tika vērtēti pret sākotnējā kartē dokumentēto; un publiskā datubāze rada reālu mācību datu pārklāšanās risku, ko arī paši autori raksturo kā iespējamu augšējo robežu saviem rezultātiem. Patiesais jaunums ir aģents, kas rīkojas visā darbplūsmā, nevis tikai atbild uz izolētiem jautājumiem. Autori skaidri norāda, ka vispārināmība, drošība un pārvaldība vēl jāpārbauda prospektīvos reālās pasaules pētījumos. Tāpēc pareizais lasījums ir: iespaidīga spēju demonstrācija, nevis pierādījums, ka MI diagnosticē labāk par ārstiem, un nevis sistēma, kas jau gatava īstai klīnikai.

Bez pārspīlējumiem

Ko pētījums parāda: Uz valodas modeļa balstīts aģents MIRA spēj izolētā EHR vidē izvadīt visu klīnisko gadījumu no sākuma līdz beigām — anamnēzi, testus, diagnozi un rīkojumus — un 574 retrospektīvu gadījumu etalonā ar astoņām diagnozēm sasniedza augstāku diagnostisko precizitāti nekā ārsti (88,9% kopumā; 87,8% pret 78,1% salīdzinājumā ar sertificētiem ārstiem), neuzrādot augstas smaguma pakāpes medikamentu kļūdas.

Kas ir ticams, bet nav pierādīts: Ka šī spēja pārvēršas ieguvumā īstiem, nediferencētiem pacientiem; ka precizitātes pārsvars atspoguļo labāku spriešanu, nevis vairāk nozīmētu testu, sakritību ar vēsturisko karti vai publisku datu iegaumēšanu.

Ko tas neparāda: Ka MIRA darbojas ārpus astoņām diagnozēm; ka to ir droši ieviest; ka tā godīgā salīdzinājumā ar identiski informētiem ārstiem pārspēj cilvēkus; ka tā aizstāj klīnicistus; vai ka rezultātos nav mācību datu pārklāšanās.

Galvenie ierobežojumi: Retrospektīvs, simulēts un tikai tekstā balstīts vērtējums; astoņas iepriekš izvēlētas diagnozes; informācijas asimetrija (daudz vairāk asins analīžu — ap 51% no pieejamajiem analītiem pret 28%, galvenokārt lēti asins testi, nevis attēldiagnostika); ārstēšanas iznākumi daļēji vērtēti pret vēsturisko karti; un publisks etalons MIMIC-IV, ko valodas modelis varēja redzēt apmācībā — paši autori to atzīmē kā iespējamu snieguma augšējo robežu.

Cik lielai pārliecībai vajadzētu būt vispārīgam lasītājam? Augstai, ka MIRA ir reāla un jauna spēju demonstrācija — aģents, kas *rīkojas* visā pacienta kartē, nevis tikai tērzē. Zemai attiecībā uz apgalvojumu, ka tā ir *labāks diagnosts par ārstu*: šis secinājums aprobežojas ar vienu retrospektīvu etalonu un ir sajaukts ar atšķirīgu testu izmantošanu, vērtēšanu pret vēsturisko karti un iespējamu datu pārklāšanos. Pašu autoru galvenais secinājums ir pareizais — pirms tam ir nozīme pacientu aprūpē, vajadzīgs prospektīvs pētījums reālajā pasaulē.

Avoti

Ferber et al., Towards autonomous medical artificial intelligence agents, Nature (2026)
https://doi.org/10.1038/s41586-026-10675-5

PubMed 42310457 — peer-reviewed abstract
https://pubmed.ncbi.nlm.nih.gov/42310457/

Science Media Centre — expert reaction to MIRA and AMIE (2026)
https://www.sciencemediacentre.org/expert-reaction-to-presentation-of-two-new-medical-ai-models-for-

News-Medical (2026) — illustrates the 'outperforms doctors' framing
https://www.news-medical.net/news/20260621/Autonomous-medical-AI-outperforms-doctors-in-simulated-EH
aspx