



The CLEAN PAPER

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

एक AI एजेंट ने पूरामरीज-ममलाखुद संभाला— लेकिन सम्युलेटर में, पुरमे रकिण्डोपर

MIRA एक अलग तरह की चिकित्सकीय AI है: एक सवाल का जवाब देने के बजाय यह नकली अस्पताल रकिण्डो के भीतर पूराममलाचलती है — इतिहास लेती है, जॉब्स मंगती और पढ़ती है, नदिम तक पहुँचती है और आदेश लिखती है। सखजनिक डेटासे के 574 पुरमे ममलो और पहले से चुने गए आठ नदिमोमें लेखकोके अनुसार इसकी नदिम-सटीकता चिकित्सकोसे अधिक रही और इसके अधिकश नरिणय दशिमरिदेशोके अनुरूप तथादवमसुरक्षाकी दृष्टिसे उचित थे। लेकिन हर सीमामहत्वपूर्ण है: यह केवल पछ-आधारित सैडबैक्स में पुरमे रकिण्डोपर चली इसकी बढ़त खसकर उन रोगोमें थीजनिके जॉब-नतीजे सफ थे; इसने डॉक्टरोसे लगभग दोगुनी रक्त-जॉब्स मंगई; और कई परणिमोकोमूल चार्ट में दर्ज उत्तर के आधार पर अंक दिए गए। असली प्रगतियह है कि एक एजेंट पूरे कख्यप्रवह में कखरवई कर सकता है — यह प्रमाण नहीं कि मशिन अब डॉक्टर से बेहतर नदिम करती है। लेखकोके अनुसार समस्यकरण, सुरक्षा और शसन के लिए अभीवस्तवकि दुनियामें भसी अध्ययन जरूरी हैं।

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

कसिसवल काजवस देनाऔर पूराममलासंभलनाएक हीवस नही है

चकित्सकीय AI कीअधकिंश कहमयिँऐसे मॉडल के बारे में होतीहै जोजवस देताहै। आप उसे कसिमरीज कासंक्षपित वविरण यापरिक्षाकासवल देते हैं, वह नदिस यासलह काएक पैरामरफ लैटताहै और अच्छास्कोर कर लेताहै। यह उपयोगी है, लेकिन दयरासीमति है: जैसे कई तेज सलहकर जोमरीज कीपूरीफइल कोकभीहय हीन लगाए।

MIRA कोकुछ अलग करने के लिए बनयागयाहै, और असलीखबर यहीअंतर है। एक सवल काजवस देने के बजय यह पूरे ममले पर कम करताहै: मरीज कारकिंड पढ़ताहै, तय करताहै कि और कैम-सीजमकरीचहए, प्रयोगशाला इमेजि और सूक्ष्मजीव-वज्जम जँवें मंगताहै, परणिस पढ़ताहै, संभवति नदिसोकीसूचीकोसंकराकरताहै और फरि आगे के आदेश लखिताहै — दवाएँ, भर्तीयासर्जरीके लिए रेफरल। यह कसिइंसम कोकमीकरने के लिए केवल मुक्त-पठ सलह नही बनता; यह इलेक्ट्रॉनिक स्वस्थय रकिंड के अंदर क्रयिँ करताहै, जैसे कई चकित्सक करताहै।

यह सचमुच एक कदम आगे है: सलह देने वलीप्रणलीसे पूरे कायप्रवह में कम करने वलीप्रणलीतक। और यहीवह तरह कीप्रगतिहै जसि खबरोमें अक्सर समेटकर “AI डॉक्टरोसे बेहतर” बनादियाजताहै। इसलिए यह सफ़ रखनाज़ूरीहै कि वस्तव में क्यापरखागयाऔर कसि परस्थितिमें।

एक वक्षय में: MIRA ने पूरे ममले अपने आप चलाए और इस बेंचमार्क पर नैदमकि सटीकतामें चकित्सकोसे बेहतर स्कोर किया— लेकिन यह सैंडबॉक्स में, पुरमे रकिंड पर, पहले से चुने गए आठ नदिसोके भीतर और केवल पठ के मध्यम से हुआ; उसकीबढ़त काबड़ाहसिसाउन स्थितियोंमें आयाजहाँजँव-परणिस बहुत स्पष्ट थे। प्रगति वस्तवकि है। लेकिन स्कोरबोर्ड क्लिनिकि नहीहोता।

MIRA showed a workflow capability, not a clinic verdict

A sandboxed EHR lets an agent act through a whole retrospective case. It is still not a real patient trial.

DEMONSTRATED

- end-to-end case work in a sandboxed EHR
- history, tests, differential diagnosis, orders
- 574 retrospective MIMIC-IV cases
- 8 pre-selected diagnoses
- higher diagnostic accuracy on this benchmark

NOT DEMONSTRATED

- real patients or live hospital deployment
- conditions beyond the eight-target set
- an equally informed head-to-head comparison
- freedom from public-data contamination
- safety and governance for clinical use

A capability shown in a sandbox -- not a diagnostician proven in a clinic.

The figure separates the workflow result from the deployment claim.

MIRA ने सैंडबॉक्स वले EHR के भीतर शुरू से अंत तक पूराकायप्रवह चलमे कीक्षमतादखिई। उसने वस्तवकि क्लिनिकल तैनी व्यपक नैदमकि दयरायासमम जमकरीके सय डॉक्टरोके वरिद्ध नषिपक्ष आमने-समने तुलनानहीदखिई।

लेखकोने क्याकिया

MIRA — Medical Intelligence for Reasoning and Action — OpenAI के मॉडलोपर बनाएक स्वयत्त एजेंट है। बातचीत और क्रियाओं वलाहसिसा GPT-4o पर चलता है, जबकि अलग योजनाचरण OpenAI के o1 रीजनगि मॉडल का उपयोग करता है। दमोएक विशेष रूप से बनाए गए, ममकोके अनुरूप ढँचे के भीतर कम करते हैं। यह HL7 FHIR पर आधारित है — वही डेटाममक जिसका उपयोग वास्तविक अस्पताल स्वस्थ रिकॉर्ड के आदान-प्रदान के लिए करते हैं — और इसमें 80,000 से अधिक संभवित क्लिनिकल क्रियाओं का उपयोग-संग्रह है। इस सैंडबॉक्स में MIRA मरीज का इतिहास देख सकता है, प्रयोगशाला इमेजिंग और सूक्ष्मजीव-वर्जित जैव मंगा और समझ सकता है, विभिन्न नदिम बना सकता है और उपचार-योजना तैयार कर सकता है — दवाएँ लिखने से लेकर सर्जरी तक करने और भर्ती की योजना बनाने तक। एक विवरण शुरू में ही ध्यान देने योग्य है: MIRA के उत्तरोक्तों को देने वाले स्वचालित “नरिणक” स्वयं GPT-4o थे — यन्नीउसीमॉडल परिवार से जैसी परखाजार था। एक सहकर्मिसमीक्षक ने इस पर आपत्त उठाई, जिसके बाद लेखकोने प्रमणित चिकित्सक की समीक्षा भी जेड़ी।

परिक्षण-समग्र MIMIC-IV से आई, जो स्वास्थ्यनिकी रूप से उपलब्ध और पहचान-रहित किया गया एक बड़ा EHR डेटाबेस है। Beth Israel Deaconess Medical Center में 2008 से 2019 के बीच इलाज पाए लगभग 300,000 मरीजों में से लेखकोने 574 मामले चुने, जो आठ लक्ष्य नदिमों में फैले थे — इनमें पेट से जुड़ी बीमारियाँ और आंतरिक चिकित्सा की आपत स्थितियाँ थीं जैसे appendicitis, pancreatitis, pneumonia और urinary-tract infection। हर मामले में MIRA समिति जनकारी से शुरू करता था और चरण-दर-चरण तय करता था कि आगे क्या करना है। संरचना वास्तविक चिकित्सकीय जैसा जैसी थी लेकिन वह ऐसे ऐतिहासिक रिकॉर्ड पर चल रही थी जिसका वास्तविक अंत पहले से फ़ैल में मौजूद था।

फिर उन्ही मामले पर उसके प्रदर्शन की तुलना चिकित्सकों से की गई।

सैंडबॉक्स वाले EHR में “स्वयत्त एजेंट” का वास्तव में क्या अर्थ है

यहाँ तक शब्द बहुत कम कर रहे हैं, इसलिए उन्हें अलग-अलग समझना उपयोगी है।

एजेंट का अर्थ है कि प्रणाली केवल किसी प्रम्प्ट का उत्तर देने वाला चैटबॉट नहीं है। यह एक चक्र चलती है: मौजूदा स्थिति देखो, कोई क्रिया चुनो — यह जैसा मंगाओ, वह इतिहास पढ़ो — परिणाम देखो फिर अगली क्रिया चुनो और यह क्रम नदिम तथा योजना तक चलता है। “बुद्धिमत्ता” भ्रामक देता है; एजेंसी उसके चारों ओर का ढँचा देता है, जो मॉडल के पष्ठ को अनुमत EHR क्रियाओं में बदलता है और परिणाम फिर मॉडल को लौटाता है।

सैंडबॉक्स वाला EHR किसी सक्रिय अस्पताल प्रणाली का नाम नहीं है, बल्कि मेडिकल-रिकॉर्ड प्रणाली की एक समियुलेटेड और अलग-थलग प्रतिलिपि है। MIRA कई जैसा “मंगा” सकता है और उस मरीज को वास्तविक जीवन में मिला परिणाम प्राप्त कर सकता है, क्योंकि मिला ऐतिहासिक है और उत्तर पहले से रिकॉर्ड में दर्ज है। उसकी कोई भी क्रिया किसी वास्तविक मरीज या क्लिनिक को प्रभावित नहीं करती।

स्वयत्त का अर्थ है कि वह सैंडबॉक्स के अंदर बना किसी ईसम को भी में शामिल किए मिला शुरू से अंत तक पूरा करता है। इसका अर्थ यह नहीं कि उसे वास्तविक मरीजों को बर्तानगरिनी संभलने के लिए छोड़ दिया गया था। ये दो बहुत अलग दम हैं, और इस अध्ययन ने केवल पहले को परखा।

सैंडबॉक्स के बारे में एक और बात आसानी से गलत समझी जा सकती है। MIRA अपनी पसंद की कोई भी जैसा मंगा सकता है, लेकिन प्रणाली परिणाम तभी लौटा सकती है जब वह जैसा उस मरीज के लिए वास्तविक जीवन में वास्तव में की गई थी। यदि वह ऐसा रक्त-पैनल माँगा है जो मरीज को मिला था तो वास्तविक मम लौटते हैं; यदि ऐसा स्कैन माँगा है जो कभी किया ही नहीं गया तो उत्तर मिला है “N/A — could not be performed”, कई गढ़ी हुई संख्यान ही। इतिहास भी इसी तरह कम करता है: रिकॉर्ड में माँगी गई जनकारी न हो तो समियुलेटेड मरीज कहता है कि उसे नहीं पता। इसलिए MIRA कई ऐसी नरिणक परिक्षण जहुई ढंग से नहीं बुला सकता जो असल में कभी हुआ ही नहीं उसे उसी समग्र के भीतर कम करना पड़ता है जो वास्तविक जैसा-प्रक्रिया में मौजूद थी।

यह अंतर इसलिए महत्वपूर्ण है क्योंकि सुरक्षितों में इस्तेमाल हमें वलाशब्द — “स्वयत्त” — सबसे आसानी से

उसके दूसरे, कहीअधिक व्यक्त अर्थ में पढ़ लिया जाता है।

उन्हें क्या मिला

बेंचमार्क पर MIRA नैदानिक सटीकता में चिकित्सकों से आगे रहा— लेकिन इस सुखीके पीछे तीन अलग संख्याएँ हैं, जिनमें अलग रखना ज़रूरी है।

सभी 574 मामले में अकेले MIRA ने 88.9% बार सही नदिश बताया। स्कोर MIMIC-IV में दर्ज वास्तविक डिसचार्ज नदिश के वरिद्ध किया गया। इसमें सही मुकामले में डॉक्टरों ने सभी 574 मामले नहीं किए; उन्होंने एक सप्ताह 311-मामलों के उपसमूह पर काम किया और उनके पास वही रिकॉर्ड तथा उपकरण थे जो MIRA को उपलब्ध थे। इन्हीं 311 मामलों में MIRA का स्कोर 87.8% था। उसकी तुलना अलग-अलग भर्ती किए गए दो समूहों से की गई। पहला वर्ग 7 से 11 वर्ष के अनुभव वाले चार प्रमणति चिकित्सक, जिनका स्कोर 78.1% रहा। दूसरा मशरूति अनुभव वाला समूह था मुख्यतः जूनियर रेजिडेंट और दो विशेषज्ञ, जो वास्तविक आपत विभाग की स्टफिंग के अधिक करीब था। उसका स्कोर 71.1% था। मामले और उपकरण समान थे — नतीजे में AI पहले, अनुभवी डॉक्टर दूसरे और जूनियर-प्रधान टीम तीसरे स्थान पर रही। शोधपत्र की अपनी सन्नधन भना है कि आठो बीमारियों में MIRA चिकित्सकों के “लगतार बरबरा रहा और अक्सर उनसे आगे निकला।”

आगे के नरिणय भी कमीज बूत रहे: अधिकांश दशमरिदेशों के अनुरूप, दवाओं के लहिज से सुरक्षति और भर्ती के नरिणयों में उपयुक्त। लेखक पाँच सुरक्षा श्रेणियों — दवा अंतर्क्रिया, गुरद के अनुसार खुरक, एलर्जी, QT जोखिम और opioid prescribing — में किसी उच्च-गंभीरता वाली दवा त्रुटि की रिपोर्ट नहीं करते और 468 में से 467 prescriptions को सही बतते हैं। सध हीवे सफ़ लखिते हैं कि प्रणाली ने “100% reliability” हसल नहीं की। सधे पढ़ें तो यह प्रभवशली परिणम है: पूरा मामला चलने वाला एजेंट नदिश के स्कोर में डॉक्टरों से आगे निकला।

लेकिन जीत की प्रकृति भी उतनी ही महत्वपूर्ण है जति नीजीत। शोधपत्र पढ़ने वाले स्वतंत्र विशेषज्ञों ने बताया कि बढ़त मुख्य रूप से कहाँ से आई। Science Media Centre के विशेषज्ञ पैनल पर University of Sheffield के Dr Wei Xing ने कहा कि AI की नैदानिक बढ़त मुख्यतः उन स्थितियों से आई जहाँ जाँच-परिणम सफ़ थे, जैसे appendicitis और pancreatitis, जिनमें नरिणयक स्कैन यालैब मम सवाल को कमीहद तक सुलझा देते हैं। Pneumonia और urinary-tract infection — आपत विभाग में आने के दोबहुत समस्य कारण — में AI और डॉक्टर दोसबसे खरब रहे और उनके बीच अंतर भी सबसे छोटा था।

दोसोपक्षों के पास उपलब्ध जमकरी में भी असमता थी और यह शोधपत्र के अपने आँकड़ों में दखिई देती है। MIRA ने लैब जाँचों पर अधिक भरोसा किया। उसने समस्य देखभाल में उपलब्ध प्रयोगशाला analytes में से लगभग 51% का उपयोग किया जबकि प्रमणति चिकित्सकों ने लगभग 28% का — प्रति मामले औसतन सत अतिरिक्त रक्त-पैरामीटर। लेखक सन्नधनी से बतते हैं कि यह डेटासेट में वास्तविक नियमति देखभाल के स्तर से कम था। इसलिए यह हर उपलब्ध जाँच मंगने की रणनीति नहीं थी। महँगी cross-sectional imaging के उपयोग में भी कोई व्यवस्थति वृद्धि नहीं दखि। फिर भी अधिक जमकरी अपने-आप नैदानिक सटीकता बढ़ा सकती है। इसलिए यह समस जमकरी वाले खलिइयों के बीच पूरी तरह समस मुकमलानही था जैसा Xing ने सधे इंगति किया। यदि किसी चिकित्सक को लगत, असुवधियादेरी की चिता किए बिना हर सस्ती रक्त-जाँच मंगने की छूट मलि, तो उसका प्रदर्शन भी काज पर बेहतर दखि सकता है।

“डॉक्टरों से आगे” का अर्थ “बेहतर डॉक्टर” नहीं है

बेंचमार्क पर जीत को ज़रूरत से ज़्यादा पढ़ना आसान है। “MIRA ने अधिक स्कोर किया” और “MIRA बेहतर है” के बीच कम से कम तीस महत्वपूर्ण बातें आती हैं।

पहली उसे किस चीज़ के विरुद्ध अंक दिए गए। University of Edinburgh की Professor Julie Jacko ने ध्यान दिलाया कि MIRA के कई प्रमुख परिणाम मूल डेटासेट में दर्ज चीज़ों के समेकित परीक्षण हैं। यही प्रणाली को दर्ज चिकित्सकीय व्यवहार को देखने पर इनमें मिल सकता है, ज़रूरी नहीं कि वह सर्वोत्तम संभव देखभाल दिखा रही हो। ऐतिहासिक चार्ट उत्तर-कुंजी बन जाते हैं। बेंचमार्क बनने का यह उचित तरीका है, लेकिन यह कष्टपूर्वक है, कभी-कभी अर्थ में सर्वोत्तम चिकित्सक नहीं। निष्पक्षता के लिए यह भी कहना चाहिए कि यह समस्या सबसे अधिक उपचार-संगत विवेक मेट्रिक्स पर लागू होती है — क्या MIRA के आदेश चार्ट से मेल खाते थे? — जबकि मुख्य नैदानिक सटीकता डिस्चार्ज नदिम के विरुद्ध मरीजों, जो केवल दस्तवेज की नकल से अधिक वास्तविक परिणामों को देखते हैं।

दूसरी पहली बातें गई जटिलता की असमता बहुत अधिक रक्त-जैव — उपलब्ध analytes कालगणना 51% बनाम 28% — अधिक सक्षम देती हैं। सटीकता के अंतर का कुछ हिसाब बेहतर तरीके से नहीं अधिक जटिलता को हल करने से आया हो सकता है।

तीसरी यह कहाँ चला। यह सैंडबॉक्स था पुराने ममलोपर, पहले से चुने गए आठ नदिमों के भीतर और केवल पठ के माध्यम से। University of Oxford के Dr Dominic Oliver के शब्दों में वास्तविक चिकित्सकीय आकलन केवल इस पर निर्भर नहीं करता कि मरीज क्या कहते हैं, बल्कि इस पर भी कि वे कैसे कहते हैं — सद्यः ही शरीरिक परीक्षण, दखिई देने वाला व्यवहार और शरीर-भाषा भी ममलोपर रखते हैं। तैयार रिकॉर्ड पढ़ने वाला पठ-आधारित एजेंट इनमें से कुछ नहीं देखता और यदि किसी मरीज की समस्या उन आठ चुने हुए नदिमों में आती ही नहीं तो यह अध्ययन उसके बारे में कुछ नहीं बता सकता।

समीक्षकों ने एक और शक्ति लेकिन महत्वपूर्ण चिंता उठाई: प्रशिक्षण-डेटा से संभावित दूषण। MIMIC-IV सार्वजनिक है और उस पर बहुत सहित्य उपलब्ध है, इसलिए खुले इंटरनेट पर प्रशिक्षित भ्रामक ने पहले से शोधपत्र, ममलो की चर्चाएँ या स्वयं डेटा के हिससे देखे हो सकते हैं। University of Sheffield के Dr Midhun Parakkal Unni ने इसे सीधे उठाया यदि कुछ उत्तर प्रशिक्षण डेटा में मौजूद थे, तो प्रदर्शन का एक हिसाब चिकित्सकीय तरीके के बजाय स्मृति हो सकता है, और स्वतंत्र पुनरुत्पादन हीदोको अलग कर सकती है। उल्लेखनीय है कि लेखक इस चिंता को टालते नहीं हैं। वे लिखते हैं कि उनके परिणमों को सन्नधनी से “संभावित ऊपरी सीमा” माना जा सकता है और वे दूसरे सार्वजनिक ममलोपर समन्वयकरण को अधिक आँक सकते हैं — यही शोधपत्र अपनी ही सुरक्षा पर समालोचना है।

इनमें से कोई बात MIRA को कम दलिचस्प नहीं बनाती। इससे केवल सही व्याख्या अधिक संतुलित होती है: पुराने ममलोपले एक बेंचमार्क पर प्रारंभिक संभव करने वाला एजेंट सफ़ा जैव-परिणाम वाली बीमारियों में डॉक्टरों से आगे निकला — कुछ हद तक अधिक जैव मंगल और कुछ हद तक दर्ज चार्ट से मेल खाते निर्णय लेकर। यह वास्तविक क्षमता का प्रदर्शन है, इस बात का फैसला नहीं कि मशीनें अब डॉक्टरों से बेहतर नदिम करती हैं।

यह क्या संभव नहीं करता

- यह नहीं दिखाता कि MIRA वास्तविक मरीजों पर कम करता है। हर ममला पुराना और सम्युलेटेड था किसी वास्तविक मरीज का इलाज MIRA ने नहीं संभाला।
- यह नहीं दिखाता कि यह आठ नदिमों के बहर भी कम करता है। पहले से चुने गए समूह से बहर की स्थितियाँ — यही चिकित्सा की अधिक जटिल और अस्पष्ट वास्तविकता — पर खीही नहीं गई।
- यह नहीं दिखाता कि इसे तैनात करना सुरक्षित है। लेखक स्वयं लिखते हैं कि समन्वयकरण, सुरक्षा और शासन के लिए अभी भविष्य-दर्शी वास्तविक दुनिया के अध्ययन चाहिए।

- यह डॉक्टरों के साथ पूरी तरह नष्टिपक्ष आमने-समने तुलनास्थिति नहीं करता। MIRA ने बहुत अधिक रक्त-जाँचों का उपयोग किया— उपलब्ध analytes कालगभग 51%, डॉक्टरों के 28% के मुकामले — और कुछ उपचार-संबंधी परीक्षण सर्वोत्तम देखभाल के स्वतंत्र ममक के बजाय ऐतिहासिक चार्ट से आंशिक रूप से मलिन करके अंकित किए गए।
- यह नहीं देखितता कि परीक्षण स्मृति के प्रभाव से मुक्त है। सर्वजनिक MIMIC-IV डेटाबेस के प्रशिक्षण-समग्री से ओवरलैप कर सकता है; इसे स्वतंत्र पुनरुत्पत्ति से ही बहर किया जा सकता है।
- इसका अर्थ “AI डॉक्टरों की जगह लेता है” नहीं है। यह ऐसानी-सहज तंत्र है जो रिकॉर्ड के भीतर क्रियाएँ कर सकता है; समीक्षकों का समग्र नष्टिपक्ष है कि वस्तु-वस्तु उपयोग चिकित्सकों के साथ सझेदारी में होगा जहाँ अधिक डॉक्टर के पास रहेगा और वह सही जमकरी भी उपलब्ध होगी जो पछ-रिकॉर्ड में नहीं आती।

सक्ष्म कतिनामजबूत है?

दो को दोहसिसोमें बाँटना चाहिए, क्योंकि दोनों के लिए सक्ष्म कीतकत बहुत अलग है।

क्षमता के प्रदर्शन के रूप में — यही यह देखने के लिए कि भिन्नमंडल एजेंट EHR सैंडबॉक्स के भीतर पूरे चिकित्सकीय मामले को इतिहास, जाँच, नदिम और आदेशों की कड़ी में शुरू से अंत तक चला सकता है — कम सचमुच नया और कभी विश्वसनीय है। यही भाग वस्तु-वस्तु प्रगति है और इसी पर सबसे अधिक ध्यान देना चाहिए।

श्रेष्ठता के दो रूप में — यही यह कहना कि MIRA चिकित्सकों से बेहतर है — सक्ष्म सीमा है और समझने से पढ़ा जमा चाहिए। यह एक खस पुराने ममलोवले बेंचमार्क, आठ नदिमों जमकरी की असमता — अधिक जाँचें — ऐतिहासिक चार्ट से बनी उत्तर-कुंजी और प्रशिक्षण-डेटादूषण की वस्तु-वस्तु संभ्रमना के भीतर लगू होता है। इतना कहना उचित है कि “इस बेंचमार्क पर एजेंट ने प्रभ्रमशली प्रदर्शन किया” इससे यह कहना उचित नहीं हो जाता कि “यह डॉक्टर से बेहतर नदिम करता है”, और लेखक भी दूसरा दमन नहीं करते।

सबसे उपयोगी दृष्टिकोण न “AI ने डॉक्टरों को हरा दिया” है, न “यह सिर्फ खलौता है।” बेहतर पढ़ाई यह है: चिकित्सकीय AI का एक नया प्रकार — जो केवल सवसोका उत्तर नहीं देता बल्कि रिकॉर्ड के भीतर क्रियाएँ करता है — एक कठिन पुराने ममलोवले बेंचमार्क पर अच्छा चला। अब उसे उस जगह खुद को समिति करना है जहाँ अभी परीक्षण नहीं हुआ: वस्तु-वस्तु, पहले से वर्गीकृत न किए गए मरीजों पर, भविष्य-दर्शी अध्ययन में और उचित नगिरी तथा शासन के तहत।

यह क्यों मगने रखता है

पछिले कुछ वर्षों में चिकित्सकीय AI का दिलचस्प सवल चुपचाप बदल गया है। पहले सवल था “क्या मंडल सही नदिम कर सकता है?” — और अधिक सफ़-सुथरे परीक्षण-समूहों पर मंडल बार-बार हाँकते रहे। MIRA इससे कठिन सवल की ओर बदल दखिता है: “क्या मंडल पूरा कम कर सकता है — जमकरी जुटा सकता है, जाँच मंगा सकता है, परीक्षण समझ सकता है, फैसला कर सकता है और कहरवाई कर सकता है — पूरे कहरप्रवह में?” यह अधिक उपयोगी और अधिक ईममदार सवल है, क्योंकि कठिन और जोखिम दोनों कहरवाई में ही समने आते हैं।

इसलिए परीक्षण को कैसे पेश किया जाता है, यह मगने रखता है। सुखी वलासहज नष्टिपक्ष — AI डॉक्टरों से बेहतर — अध्ययन के सबसे कम नए और सबसे कम समर्थित हसिसे की ओर ध्यान खींचता है। वस्तु-वस्तु में नया भाग छेदालेकिन अधिक महत्वपूर्ण है: ऐसा एजेंट जो पूरे रिकॉर्ड में आगे बढ़ सकता है। यदि यह क्षमता वस्तु-वस्तु दुनिया में भीटकती है, तो सबसे पहले यह कहरप्रवह बदलेगी यह नहीं कि उसका प्रभ्रम है। डॉक्टर गमब नहीं होता, जाँच मंगाना, परीक्षण पढ़ना और बक्रीबचे कमोकापछाकरना — यही कहरप्रवह — वह जगह है जहाँ ऐसा उपयोग सबसे पहले उतरता है।

और इससे प्रमाण काबेझ सहीदिशामें जाताहै। पुरमे ममलोके लीडरबोर्ड पर जीत भविष्य-दर्शीपरीक्षण करने काकारण है, उसकावकिलप नही। लेखक भीयहीकहते हैं। MIRA कोईममदरीसे पढ़ें तोयह अगले अध्ययन कानमित्रण है — उसी ममक पर जसिकीचकित्साकोपहले से समझ है: मरीजोपर ममें, केवल बेंचमार्क पर नही।

संक्षेप में

MIRA एक स्वयत्त AI एजेंट है जोसैंडबॉक्स वाले इलेक्ट्रॉनिक स्वस्थ रिकॉर्ड के भीतर कम करताहै। यह इतिहास ले सकताहै, प्रयोगशाला इमेजिंग और सूक्ष्मजीव-वर्जित जैव मंगाऔर समझ सकताहै, नदिम तक पहुँच सकताहै और उपचार-योजनालिख सकताहै। सवजनिक MIMIC-IV डेटासे के 574 पुरमे ममलोऔर पहले से चुने गए आठ नदिमोपर इसे परखने पर यह नैदमिक सटीकतामें चकित्सकोसे आगे रहा— कुल 88.9%; सीधे मुकबले में 87.8% बनम 78.1% — और इसके अधकिंश नरिणय दशिमरिदेशोके अनुरूप तथादवधों के लहिज से सुरक्षति थे। लेकिन मूल्यंकन केवल पठ पर आधारति पुरमे रिकॉर्ड कीसमिलेशन थी। बढत काबड़ाहिसासफ़ जैव-परणिम वालीबीमरियोमें आया। MIRA ने डॉक्टरो कीतुलनामें बहुत अधिक रक्त-जैवोकाउपयोग किया— उपलब्ध analytes कालगभग 51% बनम 28%; कुछ उपचार-संबंधी नतीजोकोमूल चार्ट में दर्ज चीजोके वरिद्ध अंकति कियोगया और सवजनिक डेटासेट प्रशिक्षण-डेटादूषण कावस्तवकि जोखिम पैदाकरताहै, जसि लेखक स्वयं संभवति ऊपरीसीमामते हैं। असलीप्रगतियह है कि एजेंट अलग-अलग सवलो काजवम देने के बजय पूरे कायप्रवह में कहरवई कर सकताहै। समस्यकरण, सुरक्षाऔर शसन के लिए अभीवस्तवकि दुनियामें भविष्य-दर्शीअध्ययन चाहिए। इसलिए सहीनषिकर्ष है: प्रभमशलीक्षमताकाप्रदर्शन, न कि यह प्रमाण कि AI डॉक्टरोसे बेहतर नदिम करताहै, और न हीऐसातंत्र जोआज वस्तवकि क्लिनिकि के लिए तैयार हो।

बनाबढ़ाचढ़कर जैव

शेषपत्र क्यादखिताहै: भ्रममंडल एजेंट MIRA सैंडबॉक्स वाले EHR के भीतर पूराचकित्सकीय ममलाशुरू से अंत तक चला सकताहै — इतिहास, जैव, नदिम और आदेश — और आठ नदिमोवले 574 पुरमे ममलोके बेंचमार्क पर नैदमिक सटीकतामें चकित्सकोसे आगे रहा कुल 88.9%, और प्रमाणति चकित्सकोके वरिद्ध सीधे मुकबले में 87.8% बनम 78.1%, जबकि कई उच्च-गंभीरतावालीदवात्रुटि दर्ज नहीहुई।

क्यासंभव है, पर सन्नति नही कियहीक्षमतावस्तवकि, पहले से वर्गकृत न किए गए मरीजोके लिए लम में बदलेगी और यह कसिटीकताकीबढत वस्तव में बेहतर तरक से आई, न कि अधिक जैव मंगमे, ऐतिहासिक चार्ट से मेल यासवजनिक डेटाकोपहले देखने से।

यह क्यानहीदखिता कि MIRA अपने आठ नदिमोके बहर कम करताहै; कि इसे वस्तवकि क्लिनिकि में तैनात करना सुरक्षति है; कि समम जमकरीवले नषिपक्ष मुकबले में यह डॉक्टरोसे आगे रहताहै; कि यह चकित्सकोकीजगह ले सकताहै; याकि परणिम प्रशिक्षण-डेटादूषण से मुक्त है।

मुख्य सीमाएँ: पुरमे ममलोपर आधारति समियुलेटेड, केवल-पठ मूल्यंकन; पहले से चुने गए आठ नदिम; जमकरीमें असममता — MIRA ने उपलब्ध analytes कालगभग 51% इस्तेमाल किया। डॉक्टरोने लगभग 28%, मुख्यतः सस्ती रक्त-जैवोमें, इमेजिंग में नही। कुछ उपचार-परणिम ऐतिहासिक चार्ट से मलिम करके अंकति किए गए; और सवजनिक MIMIC-IV बेंचमार्क कोभ्रममंडल ने प्रशिक्षण के दौरम देखाहोसकताहै, जसि लेखक स्वयं प्रदर्शन कीसंभवति ऊपरी सीमामते हैं।

समस्य पठक कोकतिनाभरोसारखनाचहिए? इस वक्त पर उच्च कि MIRA वस्तवकि और नई क्षमतादखिताहै — ऐसा एजेंट जोकेवल चैटबॉट कीतरह जवम देने के बजय रिकॉर्ड के भीतर कहरवई करताहै। इस दमे पर कम कविह चकित्सक

से बेहतर नदिमकर्ता है: वह नष्टिर्ष एक पुरमे ममलोवले बेंचमाक तक सीमिति है और अधिक ज़ाँवे मंगमे, चार्ट के वरिद्ध स्कोरिंग तथासंभवति डेटादूषण से प्रभवति होसकता है। लेखक स्वयं सहीअगलाकदम बतते हैं — मरीजोंकी देखभाल के लिए इसकाअर्थ नकिलने से पहले भविष्य-दर्शी वास्तवकि दुनियाके अध्ययन चहिए।

स्रोत

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>