The Clean Paper

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

AI agent tokko patient cases guutuu ofumaan hojjete — simulator keessatti, past records irratti

MIRA medical AI gosa haaraa: gaaffii tokko deebisuu irra simulated hospital record keessatti case guutuu hojjeta — history, tests order/read, diagnosis, orders. Public database retrospective cases 574, eight pre-selected diagnoses irratti authors diagnostic accuracy physicians caale fi largely guideline-concordant medication-safe decisions godhe jedhu. Garuu qualifier hundi matter: sandbox past records, text only; edge baay'een clear tests conditions irraa; blood tests doctors caalaa naannoo dachaa fayyadame; outcomes tokko tokko original chart against scored. Advance dhugaan agent whole workflow keessatti action fudhatu dha — machine doctor caalaa diagnose godha proof miti. Authors generalization, safety, governance prospective real-world studies barbaadu jedhu.

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

Gaaffii deebisuun haala guutuu gaggeessuu waliin tokko miti

Medical-AI stories hedduun moodeela *deebii kennu* irratti. Vignette yookaan exam gaaffii kennita, adda-baafannaa dhibee yookaan advice paragraph deebisa, score gaarii argata. Faayidaa qabu, garuu narrow: consultant smart chart hin tuqne.

MIRA waan adda godhuuf built dha, garaagarummaan kun news dhugaa dha. Gaaffii tokko deebisuu irra, haala guutuu hojjeta: dhukkubsataa galmee dubbisa, seenaa maal ammallee barbaadu murteessa, laaboraatoorii/imaging/microbiology qormaatawwan order godha, bu'aawwan dubbisa, differential adda-baafannaa dhibee dhiphisa, achiis orders itti aanan barreessa — qorichawwan ajajaman, admission, surgery referral. Free text namaa transcribe godhu baasu irra, electronic fayyaa galmee *keessatti* actions fudhata, clinician fakkaatee.

Kun tarkaanfii dhugaa: sirna advice kennu irraa sirna adeemsa hojii guutuu keessatti gocha fudhatu. Coverage keessatti "AI doctors caale" jedhamee flatten gochuun salphaa. Kanaaf maal, eessatti qorame sirrii ta'uun barbaachisaa.

One-sentence gosa: MIRA haalawwan guutuu ofumaan hojjete, madaallii wabii kana irratti adda baasuu dhibee sirrummaa'n physicians caale — garuu sandbox keessatti, retrospective galmeewwan irratti, diagnoses dura filataman saddeet keessatti, text qofa irraa communication godhee; edge isaa baay'een haalawwan qormaata bu'aawwan ifaa qaban irratti argame. Advance dhugaa dha. Scoreboard clinic miti.

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 sandboxed EHR keessatti end-to-end adeemsa hojii dandeettii agarsiise. haala addunyaa dhugaa kilinikaa hojiirra oolmaa, broad adda baasuu dhibee coverage yookaan doctors waliin equally informed fair

head-to-head hin agarsiifne.

Original Aurora diagram — The Clean Paper CC BY 4.0

Barreessitoonni maal godhan

MIRA — Yaalaa Intelligence for Reasoning fi Gocha — autonomous agent OpenAI moodeelota irratti built: converse/gocha kutaa **GPT-4o**, adda planning tarkaanfii OpenAI **o1** reasoning moodeela fayyadama. Custom standards-compliant bu'uura HL7 FHIR irratti built, toolbox danda'amaa kilinikaa actions 80,000 ol qabu, keessa taa'a. Sandbox keessatti MIRA dhukkubsataa seenaa pull, labs/imaging/microbiology order+interpret, differential adda-baafannaa dhibee generate, yaala plans formulate — qoricha prescribe, surgery schedule, admission karoora. Bal'ina ibsaa tokko: automated "judges" MIRA deebiiwwan score godhan ofii isaanii GPT-4o, wal-fakkaataa moodeela family tested; gamaaggama ogeessotaaer concern booda board-certified physician gamaaggama itti dabalame.

Qormaata set **MIMIC-IV**, ummataaf banaa de-identified EHR kuusaa daataa guddaa irraa. Beth Israel Deaconess Yaalaa Giddugala 2008–2019 dhukkubsattoota ~300,000 keessaa barreessitoonni **haalawwan 574**, **galma diagnoses saddeet** — abdominal pathologies/internal-medicine emergencies, appendicitis, pancreatitis, pneumonia, UTI dabalatee — filatan. Haala tokkoon tokkoon MIRA limited picture irraa jalqaba; step-by-step next gocha murteessa, dhugaa workup shape fakkaata, garuu galmee historical ending duraan on file irratti raawwata.

raawwii physicians wal-fakkaataa haalawwan irratti waliin wal bira qabame.

Maal "autonomous agent in a sandboxed EHR" actually means

Jechoonni sadii hojii baay'ee baatu.
Agent jechuun chatbot ajaja tokko deebisu miti. Loop raawwata: ammaa haala ilaali, gocha filadhu (qormaata order, seenaa gaafadhu), bu'aa argi, next gocha filadhu — adda-baafannaa dhibee/planga ga'utti. "Intelligence" afaan moodeela dha; *agency* scaffolding text moodeela gara permitted EHR operations jijjiiree bu'aawwan deebisee feed godhu dha.
Sandboxed EHR simulated, walled-off copy medical-record sirna dha, live hospital miti. MIRA qormaata "order" godhee bu'aa dhukkubsataa dhugaan argate receive gochuu danda'a, sababni haala historical fi deebii galmeeffame. Gocha isaa dhugaa dhukkubsataa/clinic hin tuqu.
Autonomous jechuun sandbox *keessatti* human-in-the-loop malee haala end-to-end xumura. Dhukkubsattoota unsupervised manage gochuuf set loose ta'e jechuu **miti**. Himannoowwan lamaan baay'ee adda; jalqabaa qofa qorame.
Sandbox irratti bal'ina ibsaa: MIRA qormaata kamuu *order* gochuu danda'a, garuu sirna bu'aa deebisuu danda'a qormaata sun dhukkubsataa sana irratti dhugaa jireenya keessatti **actually performed** yoo ta'e qofa. Dhiiga panel dhukkubsataa argate gaafate gatiiwwan dhugaa argata; scan nobody ordered gaafate "N/A — raawwatamuu hin dandeenye" argata, invented lakkoofsa miti. Seenaa illee galmee keessa hin jirre simulated dhukkubsataa "hin beeku" jedha.

> Kanaaf decisive qormaata kanaan dura hin fudhatamne summon hin danda'u; dhugaa workup keessatti waan ture qofaan hojjetti.
> Garaagarummaan kun barbaachisaa, sababni mata-duree "autonomous" unsupervised dhugaa dhukkubsataa management jedhamee dubbifamuuf salphaa.

Maal argatan

Madaallii wabii irratti **MIRA adda baasuu dhibee sirrummaa'n physicians caale** — garuu mata-duree lakkoofsota sadii wal-makuuf salphaa.

Haalawwan **574** hunda keessatti MIRA adda-baafannaa dhibee sirrii **88.9%**, galmeeffame discharge adda-baafannaa dhibee MIMIC-IV waliin scored. Namaa head-to-head keessatti doctors haalawwan 574 hunda hin hojjanne; shared **311-haala subset** wal-fakkaataa galmeewwan/meeshaalee MIRA waliin. Wal-fakkaataa 311 irratti MIRA **87.8%**. Garee senior board-certified physicians afur, experience 7–11 waggoota, **78.1%**; mixed-seniority irra caalaa junior residents + specialists lama, **71.1%**. Wal-fakkaataa haalawwan/meeshaalee: AI jalqabaa, experienced doctors lammaffaa, junior-heavy sadaffaa. Barruu wording careful: MIRA dhibeewwan saddeet keessatti "consistently equivalent to, fi often exceeded" physicians.

Downstream decisions illee irra caalaa guideline-concordant, qoricha-safe fi admission appropriate. Barreessitoonni five nageenya gareewwan keessatti high-severity qoricha dogoggorawwan hin gabaasa; qorichawwan ajajaman 467/468 sirrii — sirna "100% amanamummaa hin ga'in" jedhan. Fuula gatii irratti striking: agent guutuu haala raawwatee gabatee madaallii adda-baafannaa dhibee irratti doctors dura.

Garuu *shape* win barbaachisaa. Walaba specialists edge eessaa dhufe akeekan. Saayinsii Media Giddugala expert panel irratti Dr Wei Xing mata-duree AI sirrummaa advantage **irra caalaa haalawwan ifaa qormaata bu'aawwan qaban** — appendicitis/pancreatitis, decisive scan/lab — irraa driven jedhu. Pneumonia fi UTI, baramaa emergency rakkoowwan, keessatti AI fi doctors lamaan isa hamaa; gap isa xiqqaa.

Odeeffannoo asymmetry illee jira. MIRA lab irratti caalaa hirkate: routine kunuunsa jiru laaboraatoorii analytes keessaa ~**51%**, board-certified physicians ~**28%** — median dhiiga paaraameetiroota seven caalaa tokkoon tokkoon haala. Barreessitoonni kuusaa daataa routine-kunuunsa sadarkaa jalqabaa *gadi* akka ta'e, order-everything strategy miti, higher-cost imaging systematic daballii hin jiru jedhu. Garuu caalaa odeeffannoo itself adda baasuu dhibee sirrummaa dabala; kanaaf equally-informed players fair contest guutuu miti. Clinician cheap dhiiga qormaatawwan hunda cost/discomfort/delay malee order godhu illee barruu irratti sharper fakkaata.

Maaliif “doctors caale” jechuun “doctor caalaa gaarii” miti

Madaallii wabii win overread gochuun salphaa. “MIRA score ol” fi “MIRA caalaatti gaarii” gidduu wantoonni sadii jiru.

Tokkoffaa, **maal irratti scored.** Professor Julie Jacko key bu’aawwan tokko tokko underlying kuusaa daataa keessatti maal documented ta’e walbira qabamee defined jechuun, sirna **galmeeffame kilinikaa amala reproduce** gochuuf rewarded, optimal kunuunsa mirkaneessuu gochuuf miti. Historical chart deebii key ta’a. Madaallii wabii reasonable dha, garuu absolute correctness miti. Fairness: kun treatment-alignment safartuuwwan irratti strongest; mata-duree adda baasuu dhibee sirrummaa discharge adda-baafannaa dhibee irratti scored, documentation echo caalaa dhugaa bu’aa irraa dhihoo.

Lammaffaa, **odeeffannoo asymmetry**: dhiiga qormaatawwan baay’ee (51% vs 28%) ragaa baay’ee dha. sirrummaa gap keessaa tokko probable *bought* dha, reasoned qofa miti.

Sadaffaa, **eessatti raawwate.** Sandbox, retrospective haalawwan, diagnoses saddeet pre-selected, text qofa. Dr Dominic Oliver akka jedhu dhugaa assessment dhukkubsataa maal jedhu qofa miti, akkam akka jedhu, fiizikaalaa exam, mul’ate amala/mallattoo qaamaa dabalata — barreeffama qofa agent finished galmee dubbisu kana hin argu. Rakkoo diagnoses saddeet keessaa ala ta’e qorannoo scope ala.

Reviewer worry biraa: **daataa contamination.** MIMIC-IV ummataaf banaa fi heavily written waa’ee; internet-trained LLM barruulee/haala discussions/daataa ofii isaa arguu mala. Dr Midhun Parakkal Unni kun raawwii keessaa yaadannoo, kilinikaa reasoning miti, ta’uu mala; walaba replication qofa adda baasa. Barreessitoonni ofii bu’aawwan “danda’amaa gubbaa bound” jechuun cautiously interpret, biraa ummataaf banaa haalawwan generalization overestimate gochuu mala jedhu — barruu mata-duree ofii irratti ceiling kaa’a.

Kun MIRA interesting ta’uu hin hir’isu. Sirrii dubbisuu calibrated: retrospective madaallii wabii irratti guutuu-record agent ifaa-qormaata diagnoses irratti doctors caale — qormaatawwan baay’ee order gochuu fi galmeeffame chart waliin agree gochuu irraa kutaa tokko. Agarsiisa dandeettii dhugaa dha, machines amma doctors caalaa diagnose godhu verdict miti.

Kun maal *hin* mirkaneessu

- MIRA dhugaa dhukkubsattoota irratti hojjetu hin agarsiisu. Haalawwan hunda retrospective/simulated; dhukkubsataa tokko illee inni manage hin godhin.
- Diagnoses saddeet ala hojjetu hin agarsiisu. Medicine messy undifferentiated majority qormaata hin taane.
- Deploy nageenya qabu ta’uu hin agarsiisu. Barreessitoonni generalization, nageenya, governance prospective haala addunyaa dhugaa qorannoowwan barbaadu jedhu.
- Doctors waliin fair equally-informed head-to-head hin establish. Dhiiga qormaatawwan baay’ee caalaa (51% vs 28%), yaala bu’aawwan tokko tokko galmeeffame chart against scored.

- Memorization-free bu'aa hin agarsiisu. Ummataaf banaa MIMIC-IV leenjii overlap qabaachuu mala.
- "AI doctors bakka bu'a" miti. deeggarsa murtii gocha keessatti galmee; dhugaa fayyadama ogeessota yaalaa authority eegan waliin partnership ta'uu reviewers jedhu.

Ragaan hammam cimaa dha?

Himannaa lama addaan qoodi.

Agarsiisa dandeettii — moodeela afaanii agent EHR sandbox keessatti seenaa, qormaatawwan, adda-baafannaa dhibee, orders end-to-end haala guutuu raawwatu — novel fi reasonably convincing dha. Haaraa kutaa kun attention deserve godha.

Superiority himannaa — MIRA *physicians caalaa gaarii* — bounded dha. Addaa retrospective madaallii wabii, diagnoses saddeet, more-tests asymmetry, historical chart deebii key, danda'amaa leenjii contamination. "Agent madaallii wabii kana irratti impressive" jechuun ga'a; "doctor caalaa caalaatti gaarii diagnostician" jechuun hin ga'u, barreessitoonni illee lammaffaa himannaa hin godhan.

Faayidaa qabu stance "AI beats doctors" yookaan "toy qofa" miti. Haaraa yaalaa AI gosa *acts* keessatti galmee mannaa deebiiwwan gaaffilee hard retrospective madaallii wabii irratti gaariitti performed; amma dhugaa undifferentiated dhukkubsattoota prospectively, governance jalatti mirkaneessuu gochuu qaba.

Maaliif barbaachisaa dha

Yaalaa AI keessatti gaaffiin interesting jijjiirame. Dura "moodeela adda-baafannaa dhibee sirrii argataa?" ture; qormaata ifaa sets irratti eeyyee. MIRA gaaffii harder: "moodeela *job hojjachuu* danda'aa — gather, order, interpret, decide, act — guutuu adeemsa hojii?" Faayidaa qabu/honest caalaa, sababni acting keessa difficulty fi balaa lamaan jiru.

Kanaaf framing barbaachisaa. mata-duree *AI outperforms doctors* least novel/least supported kutaa irratti. Genuinely haaraa thing xiqqaa garuu consequential: agent guutuu galmee keessa sochoosu. Hold up yoo godhe, **adeemsa hojii** jijjiira, eenyu in charge ta'e dura. Doctor hin replace; ordering, interpreting, chasing bu'aawwan — adeemsa hojii — meeshaa kana jalqabaa landing.

Burden of ragaa sirriitti deebisee qindeessu. Retrospective leaderboard win prospective yaalii raawwachuuf sababa dha, substitute miti. Barreessitoonni illee akkas jedhu. Honest MIRA next qorannoo invitation dha — safarame on dhukkubsattoota, miti madaallii wabii.

Cuunfaa gabaabaa

MIRA autonomous AI agent sandboxed EHR keessatti seenaa fudhata, labs/imaging/microbiology order+interpret, adda-baafannaa dhibee ga'a, yaala plans barreessa. Ummataaf banaa MIMIC-IV haalawwan 574, diagnoses saddeet, irratti adda baasuu dhibee sirrummaa physicians caale (88.9% waliigalaan; 87.8% vs 78.1% head-to-head) fi irra caalaa guideline-concordant qoricha-safe decisions godhe. Garuu madaallii darbe galmeewwan fakkeessuu, text qofa; edge baay'een ifaa qormaata bu'aawwan haalawwan irraa; MIRA dhiiga qormaatawwan doctors caalaa baay'ee fayyadame (~51% analytes vs 28%); yaala bu'aawwan tokko tokko jalqabaa chart against scored; ummataaf banaa kuusaa daataa leenjii-data contamination balaa qaba — barreessitoonni danda'amaa gubbaa bound jedhu. Advance dhugaa isolated gaaffilee deebisuu irra adeemsa hojii guutuu keessatti gocha fudhatu agent dha. Generalization, nageenya, governance prospective haala addunyaa dhugaa qorannoowwan barbaadu. Impressive dandeettii demonstration dha, AI doctors caalaa diagnose godha ragaa miti, dhugaa clinic ready miti.

Sakatta'a ifaa

barruun maal agarsiisa: MIRA moodeela afaanii agent sandboxed EHR keessatti kilinikaa haala end-to-end raawwata; retrospective 574-haala/eight-adda-baafannaa dhibee madaallii wabii irratti adda baasuu dhibee sirrummaa physicians caale (88.9%; 87.8% vs 78.1% board-certified) fi high-severity qoricha dogoggorawwan hin gabaasa.

Waan amansiisaa ta'e garuu hin mirkanoofne: Dandeettii dhugaa undifferentiated dhukkubsataa faayidaa ta'uu; sirrummaa edge caalaatti gaarii reasoning irraa mannaa caalaa qormaatawwan, chart wal-simuu, memorized ummataaf banaa daataa.

Waan hin agarsiifne: Eight diagnoses ala works; deploy nageenya qabu; fair equally-informed doctor walbira qabuu; clinicians replace; contamination-free.

Daangaawwan ijoo: Retrospective simulated barreeffama qofa; diagnoses saddeet; odeeffannoo asymmetry dhiiga qormaatawwan 51% vs 28%; yaala bu'aawwan gariin historical chart; ummataaf banaa MIMIC-IV leenjii saaxilamuu danda'amaa, barreessitoonni raawwii gubbaa bound ta'uu malu jedhu.

Dubbisaan waliigalaa amanamummaa hammam qabaachuu qaba? MIRA dhugaa novel dandeettii — agent galmee keessatti *acts*, chatbot qofa miti — irratti ol-aanaa. *Doctor caalaa diagnostician* ta'uu irratti gadi-aanaa: madaallii wabii tokko fi confounds qormaatawwan/scoring/contamination. Barreessitoonni dhumarratti sirrii: dhukkubsataa kunuunsa meaning qabaachuuf prospective haala addunyaa dhugaa qorannoo barbaada.

Maddoota

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