ԱԲ գործակալն ինքնուրույն վարեց ամբողջական հիվանդների դեպքեր՝ սիմուլյացիայում, անցյալ գրառումների վրա

MIRA-ն բժշկական ԱԲ-ի նոր տեսակ է. մեկ հարցի պատասխանելու փոխարեն այն սիմուլացված հիվանդանոցային գրառման ներսում վարում է ամբողջ դեպքը՝ վերցնում պատմությունը, նշանակում և կարդում հետազոտությունները, հասնում ախտորոշման և կազմում հրահանգները: Հանրային տվյալների բազայից վերցված 574 հետահայաց դեպքերի վրա, ութ նախապես ընտրված ախտորոշումների շրջանակում, հեղինակները հայտնում են, որ այն ախտորոշման ճշտությամբ գերազանցել է բժիշկներին և հիմնականում տվել է ուղեցույցներին համապատասխան, դեղորայքային առումով անվտանգ որոշումներ: Բայց այդ նախադասության յուրաքանչյուր սահմանափակում կարևոր է. այն աշխատել է մեկուսացված միջավայրում՝ անցյալ գրառումներով և միայն տեքստով, առավելության մեծ մասը եկել է հստակ թեստերով վիճակներից, այն մոտ կրկնակի շատ արյան հետազոտություններ է օգտագործել, իսկ մի շարք արդյունքներ գնահատվել են ըստ պատմական քարտում գրանցված գործողությունների: Իրական առաջընթացը ամբողջ աշխատանքային ընթացքի միջով գործող գործակալն է, ոչ ապացույց, որ մեքենան այժմ բժիշկներից լավ է ախտորոշում: Հեղինակներն էլ ընդգծում են, որ ընդհանրացումը, անվտանգությունն ու կառավարումը դեռ պահանջում են հեռանկարային, իրական միջավայրի ուսումնասիրություններ:

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

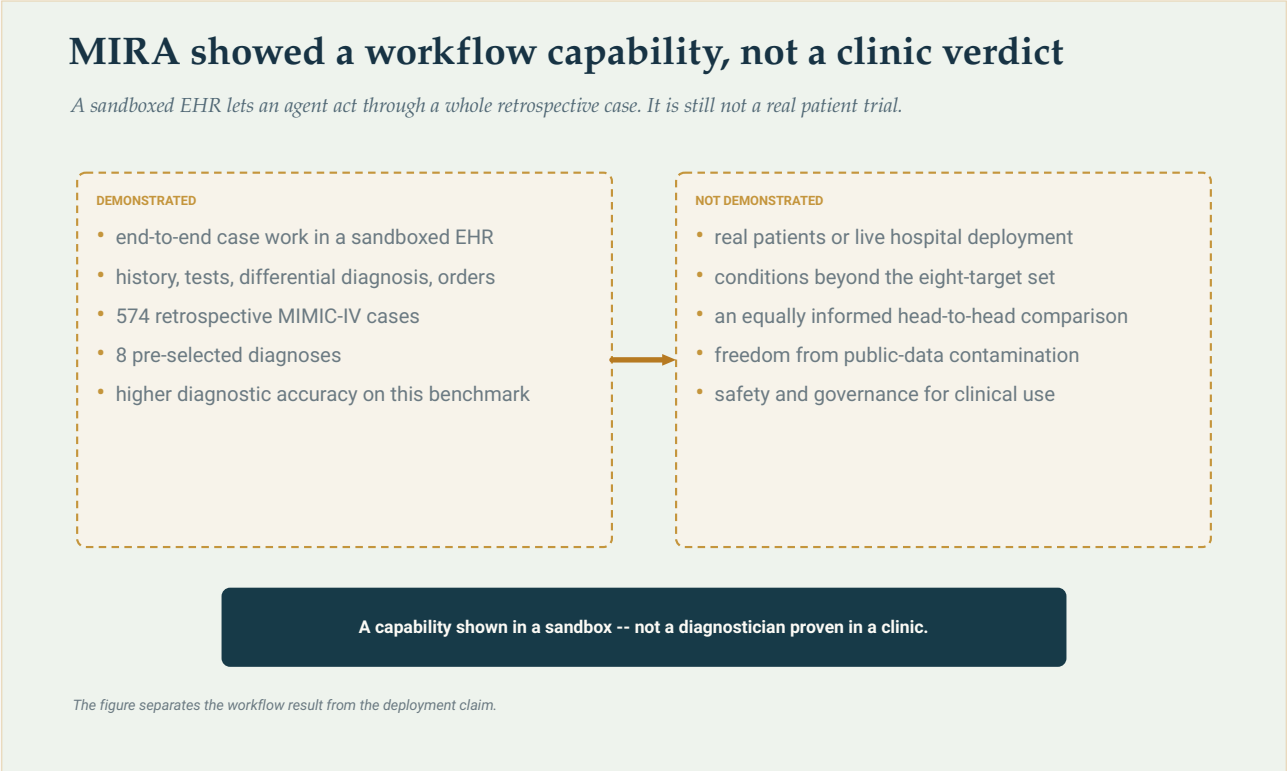
Հարցին պատասխանելը նույնը չէ, ինչ ամբողջ կլինիկական դեպքը վարելը

Բժշկական ԱԲի մասին պատմությունների մեծ մասը վերաբերում է մոդելի, որը *պատասխանում է*: Նրան տալիս են կլինիկական նկարագրություն կամ քննության հարց, իսկ այն վերադարձնում է ախտորոշում կամ խորհրդի մի պարբերություն և հաճախ բարձր գնահատական է ստանում: Սա օգտակար է, բայց սահմանափակ՝ մի տեսակ խելացի խորհրդատու, որը երբեք չի աշխատում բժշկական քարտի հետ:

MIRA-ն ստեղծվել է այլ բան անելու համար, և հենց այդ տարբերությունն է այս աշխատանքի իրական նորությունը: Մի հարցի պատասխանելու փոխարեն այն վարում է ամբողջ դեպքը. կարդում է հիվանդի գրառումները, որոշում է՝ պատմությունից դեռ ինչ է պետք պարզել, նշանակում է լաբորատոր, պատկերային և մանրէաբանական հետազոտություններ, կարդում է արդյունքները, նեղացնում տարբերակիչ ախտորոշումների շրջանակը, ապա կազմում է հաջորդ քայլերի հրահանգները՝ դեղատոմս, հոսպիտալացում կամ վիրաբուժական խորհրդատվություն: Այն գործողություններ է կատարում *էլեկտրոնային առողջապահական գրառման ներսում*՝ մոտավորապես այն ձևով, ինչպես կլինիցիստը, այլ ոչ թե պարզապես ազատ տեքստ է գրում, որ մարդը հետո փոխանցի համակարգին:

Սա իրական քայլ է՝ խորհրդատու համակարգից դեպի համակարգ, որը կարող է գործել ամբողջ աշխատանքային ընթացքի ընթացքում: Բայց հենց այսպիսի արդյունքներն են լրատվական լուսաբանման մեջ հաճախ հարթեցվում մինչև «ԱԲն գերազանցեց բժիշկներին» ձևակերպումը: Ուստի կարևոր է շատ ճշգրիտ տարբերակել՝ ինչն է փորձարկվել և որտեղ:

Մեկ նախադասությամբ. MIRA-ն ինքնուրույն վարել է ամբողջական դեպքեր և այս բենչմարքում ախտորոշման ճշտությամբ գերազանցել է բժիշկներին, բայց դա արել է մեկուսացված փորձնական միջավայրում, հետահայաց գրառումներով, նախապես ընտրված ութ ախտորոշումների շրջանակում, միայն տեքստային հաղորդակցությամբ, և առավելության մեծ մասը ստացել է այն վիճակներում, որտեղ որոշիչ թեստերի արդյունքները համեմատաբար հստակ էին: Առաջընթացը իրական է: Գնահատման աղյուսակը դեռ կլինիկա չէ:



MIRA-ն ցուցադրեց ամբողջ աշխատանքային շղթան վարելու կարողություն մեկուսացված էլեկտրոնային առողջապահական գրառման միջավայրում: Այն չցուցադրեց իրական կլինիկական ներդրում, ախտորոշումների լայն ծածկույթ կամ բժիշկների հետ հավասար տեղեկատվությամբ իրականացված արդար ուղիղ համեմատություն:

Ինչ արեցին հեղինակները

MIRA-ն՝ բժշկական Intelligence for Reasoning and Action, OpenAI-ի մոդելների վրա կառուցված ինքնավար գործակալ է: Հաղորդակցող և գործող մասը աշխատում է **GPT-4o**-ով, իսկ առանձին պլանավորման փուլում օգտագործվում է OpenAI-ի **o1** reasoning մոդելը: Դրանք տեղադրված են հատուկ, ստանդարտներին համապատասխան համակարգի մեջ, որը կառուցված է HL7 FHIR-ի վրա՝ այն տվյալային ստանդարտի, որով իրական հիվանդանոցները փոխանակում են բժշկական գրառումները: Համակարգը գործակալին տալիս է ավելի քան ութսուն հազար հնարավոր կլինիկական գործողությունների գործիքակազմ: Այդ մեկուսացված միջավայրում MIRA-ն կարող է ստանալ հիվանդի պատմությունը, նշանակել և մեկնաբանել լաբորատոր, պատկերային ու մանրէաբանական հետազոտություններ, կազմել տարբերակիչ ախտորոշումների ցանկ և ձևակերպել բուժման պլան՝ ներառյալ դեղերի նշանակումը, վիրաբուժական միջամտության պլանավորումը և հոսպիտալացումը: Մի կարևոր մանրամասն պետք է նշել հենց սկզբում. MIRA-ի պատասխանները գնահատող ավտոմատ «դատավորներն» իրենք էլ GPT-4o էին՝ նույն մոդելային ընտանիքից, որը գնահատվում էր: Գրախոսներից մեկի բարձրացրած մտահոգությունից հետո հեղինակները այս գնահատումը

լրացրել են մասնագիտական հավաստագրում ունեցող բժշկի ստուգմամբ։

Փորձարկման տվյալները վերցվել են **MIMIC-IV**-ից՝ մեծ, հանրային, ապանույնականացված էլեկտրոնային բժշկական գրառումների բազայից։ Beth Israel Deaconess բժշկական Center-ում 2008–2019 թվականներին բուժված մոտ 300 000 հիվանդների միջից հեղինակներն ընտրել են **574 դեպք**, որոնք ընդգրկում էին **ութ թիրախային ախտորոշում**՝ որովայնային ախտաբանություններ և ներքին բժշկության շտապ վիճակներ, այդ թվում՝ ապենդիցիտ, պանկրեատիտ, թոքաբորբ և միզուղիների վարակ։ Յուրաքանչյուր դեպքում MIRA-ն սկսում էր սահմանափակ տեղեկատվությունից և քայլ առ քայլ պետք է որոշեր՝ ինչ անել հաջորդը։ Կառուցվածքը նման է իրական կլինիկական հետազոտությանը, բայց այն կատարվում է պատմական գրառման վրա, որի իրական ավարտն արդեն առկա է տվյալների բազայում։

Այնուհետև նույն դեպքերի վրա MIRA-ի աշխատանքը համեմատվել է բժիշկների աշխատանքի հետ։

> **Ի՞նչ է իրականում նշանակում «ինքնավար գործակալ մեկուսացված EHR-ում»**
>
> Այստեղ երեք բառ շատ ծանրաբեռնված իմաստ ունեն, ուստի արժե դրանք բացել։
> **Գործակալ** նշանակում է, որ համակարգը պարզապես հարցին պատասխանող չաթբոտ չէ։ Այն աշխատում է օղակով. դիտում է ընթացիկ վիճակը, ընտրում գործողություն՝ օրինակ նշանակել այս հետազոտությունը կամ հարցնել այդ տվյալը, ստանում է արդյունքը, ապա ընտրում հաջորդ քայլը՝ մինչև հասնում է ախտորոշման և պլանի։ «Բանական» մասը լեզվային մոդելն է, իսկ *գործակալայինությունը* դրա շուրջ կառուցված ենթակառուցվածքն է, որը մոդելի տեքստը վերածում է թույլատրելի EHR գործողությունների և արդյունքները վերադարձնում մոդելին։
> **Մեկուսացված EHR** նշանակում է բժշկական գրառումների համակարգի սիմուլյացված, պատերով առանձնացված պատճեն, ոչ թե գործող հիվանդանոցային համակարգ։ MIRA-ն կարող է «նշանակել» հետազոտություն և ստանալ այն արդյունքը, որն այդ հիվանդն իրականում ունեցել է, քանի որ դեպքը պատմական է և պատասխանը արդեն գրանցված է։ Նրա գործողություններից ոչ մեկը չի ազդում իրական հիվանդի կամ կլինիկայի վրա։
> **Ինքնավար** նշանակում է, որ այն դեպքը սկզբից մինչև վերջ ավարտում է առանց մարդու միջամտության՝ *միայն այդ սիմուլյացիայի ներսում*։ Դա **չի** նշանակում, որ համակարգին թույլ են տվել առանց հսկողության վարել իրական հիվանդների։ Սրանք լրիվ տարբեր պնդումներ են, և փորձարկվել է միայն առաջինը։
> Մեկ այլ նրբություն էլ կա։ MIRA-ն կարող է *պահանջել* ցանկացած հետազոտություն, բայց համակարգը կարող է արդյունք վերադարձնել միայն այն դեպքում, եթե այդ հետազոտությունն **իրականում կատարվել է** տվյալ հիվանդի համար։ Եթե պահանջվի արյան այն անալիզը, որը հիվանդը ստացել

> է, վերադարձվում են իրական արժեքները: Եթե պահանջվի սկանավորում, որը երբեք չի արվել, համակարգը վերադարձնում է «N/A — չէր կարող կատարվել», ոչ թե հորինված թիվ: Հիվանդի պատմությունն էլ նույն կերպ է աշխատում. եթե գրառման մեջ MIRA-ի հարցած տվյալը չկա, սիմուլացված հիվանդը պարզապես ասում է, որ չգիտի: Այսպիսով MIRA-ն չի կարող «ստեղծել» այն վճռորոշ թեստը, որը իրական կյանքում չի կատարվել. այն սահմանափակված է իրական պատմական հետազոտության մեջ եղած նյութով:
> Այս տարբերակումը կարևոր է, որովհետև վերնագրերում ամենաշատը գերագնահատվելու վտանգ ունեցող բառը հենց «ինքնավար»-ն է:

Ինչ գտան

Բենչմարքում **MIRA-ն ախտորոշման ճշտությամբ գերազանցեց բժիշկներին**, բայց վերնագիրը թաքցնում է երեք տարբեր թիվ, որոնք հեշտ է իրար խառնել:

Բոլոր **574 դեպքերում** MIRA-ն ճիշտ ախտորոշումը տվել է **88.9%** դեպքերում՝ համեմատված MIMIC-IV-ում գրանցված վերջնական դուրսգրման ախտորոշման հետ: Սակայն բժիշկների հետ ուղիղ համեմատությունում նրանք չեն աշխատել բոլոր 574 դեպքերի վրա: Սարդիկ գնահատել են ընդհանուր **311 դեպքերի ենթաբազմությունը**, օգտագործելով նույն գրառումներն ու գործիքները, որոնք հասանելի էին MIRA-ին: Այդ նույն 311 դեպքերում MIRA-ի ճշտությունը եղել է **87.8%**: Այն համեմատվել է երկու առանձին հավաքագրված խմբերի հետ: Առաջինը **փորձառու խումբն** էր՝ մասնագիտական հավաստագրում ունեցող չորս բժիշկ, 7–11 տարվա փորձով, որոնց արդյունքը եղել է **78.1%**: Երկրորդը **տարբեր փորձառության խումբն** էր՝ հիմնականում կրտսեր ռեզիդենտներ և երկու մասնագետ, այսինքն՝ կազմով ավելի մոտ իրական շտապօգնության բաժանմունքին: Նրանց արդյունքը եղել է **71.1%**: Նույն դեպքերը, նույն գործիքները, և դասավորությունը եղել է՝ առաջինը ԱԲն, հետո փորձառու բժիշկները, ապա կրտսերներով գերակշռող խումբը: Հոդվածի զգույշ ձևակերպմամբ՝ բոլոր ութ հիվանդությունների դեպքում MIRA-ն եղել է բժիշկներին «հետևողականորեն համարժեք և հաճախ գերազանցել է»:

Հաջորդ քայլերի որոշումներն էլ ընդհանուր առմամբ պահպանել են կլինիկական ուղեցույցները և դեղորայքային անվտանգության պահանջները: Հեղինակները հինգ անվտանգության կատեգորիաներում՝ դեղերի փոխազդեցություններ, երիկամային դեղաչափեր, ալերգիաներ, QT ռիսկ և օփիոիդների նշանակում, բարձր ծանրության դեղորայքային սխալներ չեն գրանցել: Դեղատոմսերը ճիշտ են եղել 468 դեպքից 467-ում, միաժամանակ ընդգծելով, որ համակարգը «100% հուսալիության չի հասել»: Առաջին հայացքից սա տպավորիչ արդյունք է՝ գործակալ, որը վարում է ամբողջ դեպքը և ախտորոշման ցուցանիշով առաջ է անցնում բժիշկներից:

Բայց հաղթանակի *կառուցվածքը* նույնքան կարևոր է, որքան հաղթանակը: Աշխատանքը կարդացած անկախ մասնագետները նշել են, թե որտեղից է հիմնականում գալիս տարբերությունը: Science մեդիա Centre-ի փորձագիտական քննարկման մեջ

Շեֆիլդի համալսարանի դոկտոր Wei Xing-ը նշել է, որ «ԱԲն ախտորոշման ճշտությամբ գերազանցեց բժիշկներին» ընդհանուր ցուցանիշը **հիմնականում պայմանավորված էր այն վիճակներով, որտեղ թեստերը տալիս են հստակ պատասխան**, օրինակ՝ ապենդիցիտը և պանկրեատիտը, որոնց դեպքում որոշիչ պատկերային կամ լաբորատոր արդյունքը մեծապես լուծում է հարցը։ Թոքաբորբի և միզուղիների վարակի դեպքում՝ երկու շատ տարածված պատճառ, որոնցով մարդիկ դիմում են շտապ բաժանմունք, *և՛* ԱԲն, *և՛* բժիշկները ավելի վատ են աշխատել, իսկ նրանց միջև տարբերությունն ամենափոքրն է եղել։

Կար նաև տեղեկատվության օգտագործման անհամաչափություն, և այն երևում է հենց հոդվածի թվերում։ MIRA-ն ավելի լայնորեն էր օգտագործում լաբորատոր տվյալները. այն օգտվել է սովորական բուժման ժամանակ հասանելի լաբորատոր ցուցանիշների մոտ **51%**-ից, մինչդեռ փորձառու բժիշկները՝ մոտ **28%**-ից։ Մեկ դեպքի հաշվով սա միջինում յոթ լրացուցիչ արյան ցուցանիշ է։ Հեղինակները նշում են, որ սա նույնիսկ տվյալների բազայի սովորական բուժման ամբողջ ծավալից պակաս է և «ամեն ինչ պատվիրելու» ռազմավարություն չի եղել։ Նրանք նաև չեն տեսել թանկարժեք խաչաձև կտրվածքի պատկերային հետազոտությունների համակարգված աճ։ Այնուամենայնիվ, ավելի շատ տեղեկատվությունն ինքնին կարող է բարձրացնել ախտորոշման ճշտությունը։ Հետևաբար սա հավասար տեղեկատվությամբ երկու կողմերի կատարյալ համեմատություն չէ, և Xing-ը հենց այս բացը նշել է։ Բժիշկը, որը կարողանար առանց արժեքի, անհարմարության կամ ուշացման մասին մտածելու պատվիրել բոլոր էժան արյան թեստերը, նույնպես կարող էր թղթի վրա ավելի ճշգրիտ երևալ։

Ինչու «գերազանցեց բժիշկներին» չի նշանակում «ավելի լավ բժիշկ է»

Բենչմարքի հաղթանակը հեշտ է չափազանց մեկնաբանել։ «MIRA-ն ավելի բարձր միավոր ստացավ» և «MIRA-ն ավելի լավ բժիշկ է» պնդումների միջև առնվազն երեք կարևոր տարբերություն կա։

Առաջինը՝ **ինչի հետ էր համեմատվում արդյունքը**։ Էդինբուրգի համալսարանի պրոֆեսոր Julie Jacko-ն նշել է, որ MIRA-ի մի շարք հիմնական արդյունքներ սահմանված են հիմքում ընկած տվյալների բազայում փաստագրված գործողությունների համեմատությամբ Այսինքն՝ համակարգը պարգևատրվում է **գրանցված կլինիկական վարքագիծը վերարտադրելու** համար, ոչ պարտադիր այն բանի համար, որ ցույց է տալիս օպտիմալ խնամք։ Պատմական բժշկական քարտը դառնում է պատասխանների բանալի։ Սա բենչմարք կառուցելու ողջամիտ ձև է, բայց չափում է համաձայնությունը կատարվածի հետ, ոչ թե բացարձակ իմաստով լավագույն բուժումը։ Արդարության համար պետք է նշել, որ այս խնդիրը ամենից շատ վերաբերում է *բուժման համապատասխանու* չափումներին՝ արդյոք MIRA-ի հրահանգները համապատասխանե՞լ են քարտում գրանցվածին։ Գլխավոր ախտորոշիչ ճշտությունը համեմատվում է դուրսգրման ախտորոշման հետ, որը իրական արդյունքին ավելի մոտ ցուցանիշ է, քան պարզապես

փաստաթղթի կրկնությունը:

Երկրորդը՝ արդեն նշված **տեղեկատվական անհամաչափությունն** է: Արյան զգալիորեն ավելի շատ ցուցանիշների օգտագործումը՝ մոտ 51% ընդդեմ 28%-ի, նշանակում է ավելի շատ ապացույց: Ճշտության տարբերության մի մասը հավանաբար գալիս է հենց լրացուցիչ տվյալներից, ոչ միայն ավելի լավ դատողությունից:

Երրորդը՝ **որտեղ էր աշխատում համակարգը**: Սա սիմուլյացիա էր, պատմական դեպքերի վրա, ութ նախապես ընտրված ախտորոշումների շրջանակում և միայն տեքստով: Ինչպես նշել է Օքսֆորդի համալսարանի դոկտոր Dominic Oliver-ը, իրական կլինիկական գնահատումը կախված է ոչ միայն նրանից, թե ինչ է հիվանդն ասում, այլ նաև՝ ինչպես է ասում, ինչպես նաև ֆիզիկական զննումից, դիտվող վարքագծից և մարմնի լեզվից: Դրանցից ոչ մեկը չի տեսնում տեքստային գործակալը, որը կարդում է արդեն ավարտված քարտը: Իսկ եթե հիվանդի խնդիրը չի պատկանում ընտրված ութ ախտորոշումներից մեկին, այս ուսումնասիրությունն այդ դեպքի մասին պարզապես ոչինչ չի կարող ասել:

Գրախոսների բարձրացրած ավելի լուռ մտահոգությունն էլ **ուսուցման տվյալների աղտոտումն** է: MIMIC-IV-ը հանրային է և լայնորեն ուսումնասիրված, ուստի բաց ինտերնետով ուսուցված լեզվային մոդելը կարող էր արդեն տեսած լինել հոդվածներ, դեպքերի քննարկումներ կամ նույնիսկ տվյալների մի մասը: Շեֆիլդի համալսարանի դոկտոր Midhun Parakkal Unni-ն սա նշել է ուղիղ. եթե պատասխանների մի մասը եղել է ուսուցման տվյալներում, արդյունքի մի մասը կարող է լինել հիշողություն, ոչ կլինիկական դատողություն, և դա կարելի է տարբերակել միայն անկախ վերարտադրությամբ: Հատկանշական է, որ հեղինակները խնդիրը չեն անտեսում. նրանք գրում են, որ իրենց արդյունքները «կարելի է զգուշությամբ մեկնաբանել որպես հնարավոր վերին սահման» և որ դրանք «կարող են գերագնահատել այլ հանրային դեպքերի վրա ընդհանրացման ունակությունը»: Այսինքն՝ հենց հոդվածն է սահման դնում իր ամենագրավիչ թվերին:

Այս ամենը MIRA-ն պակաս հետաքրքիր չի դարձնում: Պարզապես ճիշտ ընթերցումը ավելի չափավոր է. հետահայաց բենչմարքում ամբողջ գրառման միջով գործող գործակալը բժիշկներից բարձր է գնահատվել այն ախտորոշումներում, որտեղ թեստերը հստակ էին՝ մասամբ ավելի շատ հետազոտություններ պատվիրելով և մասամբ պատմական քարտին համապատասխանելով: Սա իրական կարողության ցուցադրում է, ոչ դատավճիռ, որ մեքենաները այժմ բժիշկներից ավելի լավ են ախտորոշում:

Ինչ *չի* ապացուցում այս աշխատանքը

- Այն **չի** ցույց տալիս, որ MIRA-ն աշխատում է իրական հիվանդների վրա: Բոլոր դեպքերը հետահայաց և սիմուլացված էին, և համակարգը ոչ մի հիվանդի չի վարել:
- Այն **չի** ցույց տալիս, որ աշխատում է ութ ախտորոշումներից դուրս: Նախապես

ընտրված ցանկից դուրս գտնվող հիվանդությունները՝ բժշկության խառը և անորոշ մեծ մասը, չեն փորձարկվել։

- Այն **չի** ցույց տալիս, որ համակարգը պատրաստ է անվտանգ ներդրման։ Հեղինակներն իրենք գրում են, որ ընդհանրացումը, անվտանգությունն ու կառավարումը պետք է ստուգվեն հեռանկարային, իրական միջավայրի ուսումնասիրություններում։
- Այն **չի** հաստատում բժիշկների հետ լիովին արդար համեմատություն։ MIRA-ն օգտագործել է զգալիորեն ավելի շատ արյան ցուցանիշներ՝ հասանելի անալիտների մոտ 51%-ը ընդդեմ 28%-ի, իսկ բուժման մի շարք արդյունքներ մասամբ գնահատվել են պատմական քարտի հետ համապատասխանությամբ, ոչ թե օբյեկտիվ «լավագույն բուժման» չափանիշով։
- Այն **չի** ցույց տալիս, որ արդյունքը զերծ է հիշողությունից։ Հանրային MIMIC-IV տվյալները կարող էին համընկնել մոդելի ուսուցման տվյալների հետ, ինչը կարող է բացառել միայն անկախ կրկնությունը։
- Այն **չի** նշանակում «ԱԲն փոխարինում է բժիշկներին»։ Սա որոշումների աջակցության գործիք է, որը կարող է *գործել* գրառման ներսում։ Փորձագետների ընդհանուր գնահատմամբ՝ իրական օգտագործումը լինելու է կլինիցիստների հետ համագործակցությա որտեղ իշխանությունն ու պատասխանատվությունը մնում են մարդկանց մոտ, իսկ բժիշկները լրացնում են այն ամենը, ինչը տեքստային քարտում չկա։

Որքա՞ն ուժեղ են ապացույցները

Պնդումը պետք է բաժանել երկու մասի, որովհետև դրանց ապացույցների ուժը շատ տարբեր է։

Որպես կարողության ցուցադրում՝ որ լեզվային մոդելի գործակալը կարող է EHR սիմուլյացիայի ներսում վարել ամբողջ կլինիկական դեպքը՝ հերթականությամբ միացնելով պատմությունը, հետազոտությունները, ախտորոշումը և հրահանգները, աշխատանքը իսկապես նոր է և բավական համոզիչ։ Հենց սա է նոր մասը և ուշադրության արժանի մասը։

Որպես գերազանցության պնդում՝ որ MIRA-ն *ավելի լավ է, քան բժիշկները*, ապացույցները սահմանափակ են և պահանջում են զգույշ ընթերցում։ Արդյունքը ստացվել է մեկ կոնկրետ հետահայաց բենչմարքում, ութ ախտորոշման վրա, տեղեկատվական անհամաչափությամբ՝ ավելի շատ թեստեր, պատմական քարտից ստացված պատասխանների բանալիով և ուսուցման տվյալների հնարավոր աղտոտմամբ։ Սա բավարար է ասելու՝ «գործակալը այս բենչմարքում տպավորիչ արդյունք է տվել»։ Բավարար չէ ասելու՝ «գործակալն ավելի լավ ախտորոշող է, քան բժիշկը»։ Հեղինակներն էլ երկրորդը չեն պնդում։

Ամենաօգտակար դիրքորոշումը ոչ «ԱԲն հաղթեց բժիշկներին» է, ոչ էլ «սա պարզապես խաղալիք է»։ Ավելի ճշգրիտ ձևակերպումը սա է. բժշկական ԱԲի նոր տեսակ՝ գործակալ, որը հարցերին պատասխանելու փոխարեն *գործում է* ամբողջ գրառման ընթացքում, լավ արդյունք է տվել դժվար հետահայաց բենչմարքում և այժմ պետք է իրեն ապացուցի այնտեղ, որտեղ դեռ չի փորձարկվել՝ իրական, նախապես չընտրված

հիվանդների վրա, հեռանկարային ձևով և հստակ կառավարման ներքո:

Ինչու է սա կարևոր

Վերջին տարիներին բժշկական ԱԲի հետաքրքիր հարցը աննկատ փոխվել է: Նախկինում այն էր՝ «կարո՞ղ է մոդելը ճիշտ ախտորոշել»: Եվ մոդելները ավելի ու ավելի մաքուր թեստային հավաքածուներում հաճախ պատասխանում էին՝ այո: MIRA-ն հարցը տեղափոխում է ավելի բարդ մակարդակ՝ «կարո՞ղ է մոդելը *անել աշխատանքը*՝ հավաքել տվյալները, նշանակել, մեկնաբանել, որոշել և գործել ամբողջ ընթացքի մեջ»: Սա ավելի օգտակար և ավելի ազնիվ հարց է, որովհետև հենց գործողության մեջ են կենտրոնացած և՛ դժվարությունը, և՛ ռիսկը:

Դրա համար էլ ձևակերպումն այնքան կարևոր է: Վերնագրային բնազդը՝ *ԱԲն գերազանցեց բժիշկներին*, մատնանշում է արդյունքի ամենաքիչ նոր և ամենաքիչ հաստատված մասը: Իրապես նոր բանն ավելի սահմանափակ է, բայց ավելի հետևանքային՝ գործակալ, որը կարող է շարժվել ամբողջ բժշկական գրառման միջով: Եթե այդ կարողությունը հաստատվի, այն նախ փոխելու է **աշխատանքային ընթացքը**, շատ ավելի շուտ, քան կփոխի՝ ով է դրա պատասխանատուն: Բժիշկը չի անհետանում. նախ փոխվում են թեստեր պատվիրելը, արդյունքները հետևելը, մեկնաբանելը և հաջորդ քայլերը կազմակերպելը:

Եվ սա նաև ապացույցների պահանջը ճիշտ ուղղությամբ է տեղափոխում: Հետահայաց դեպքերի վարկանիշային հաղթանակը պատճառ է հեռանկարային փորձարկում անելու, ոչ այդ փորձարկման փոխարինողը: Հեղինակներն էլ հենց դա են ասում: MIRA-ի ազնիվ ընթերցումը հաջորդ ուսումնասիրության հրավեր է՝ այն չափանիշով, որն այս ոլորտն արդեն գիտի. չափել հիվանդների վրա, ոչ միայն բենչմարքերում:

Կարճ ամփոփում

MIRA-ն ինքնավար ԱԲ գործակալ է, որն աշխատում է մեկուսացված էլեկտրոնային առողջապահական գրառման համակարգում. այն կարող է վերցնել պատմությունը, նշանակել և մեկնաբանել լաբորատոր, պատկերային և մանրէաբանական հետազոտություններ հասնել ախտորոշման և կազմել բուժման հրահանգներ: Հանրային MIMIC-IV բազայից վերցված 574 հետահայաց դեպքերի վրա, ութ նախապես ընտրված ախտորոշումների շրջանակում, այն ախտորոշման ճշտությամբ գերազանցել է բժիշկներին՝ 88.9% ընդհանուր, իսկ ուղիղ 311-դեպքային համեմատությունում՝ 87.8% ընդդեմ փորձառու բժիշկների 78.1%-ի, և ընդհանուր առմամբ տվել է ուղեցույցներին համապատասխան ու դեղորայքային առումով անվտանգ որոշումներ: Բայց գնահատումը պատմական գրառումների սիմուլյացիա էր, միայն տեքստով: Առավելության զգալի մասը ստացվել է հստակ թեստերով վիճակներում: MIRA-ն օգտագործել է շատ ավելի շատ արյան ցուցանիշներ, քան բժիշկները՝ հասանելի անալիտների

մոտ 51%-ը ընդդեմ 28%-ի: Բուժման մի շարք արդյունքներ գնահատվել են ըստ պատմական քարտում գրանցված գործողությունների, իսկ հանրային տվյալների բազան ստեղծում է ուսուցման տվյալների աղտոտման իրական ռիսկ: Հեղինակներն իրենք այդ թվերը նկարագրում են որպես հնարավոր վերին սահման: Իրական առաջընթացը ամբողջ workflow-ի միջով գործող գործակալն է, ոչ ապացույց, որ մեքենան արդեն բժիշկներից լավ է ախտորոշում: Ընդհանրացումը, անվտանգությունն ու կառավարումը դեռ պետք է ստուգվեն հեռանկարային, իրական միջավայրի ուսումնասիրություններում:

Առանց չափազանցության

Ինչ է ցույց տալիս հոդվածը. Լեզվային մոդելի վրա հիմնված MIRA գործակալը կարող է մեկուսացված EHR-ում ամբողջ կլինիկական դեպքը վարել սկզբից մինչև վերջ՝ պատմություն, հետազոտություններ, ախտորոշում և հրահանգներ, և 574 հետահայաց դեպքերի բենչմարքում ութ ախտորոշման վրա ախտորոշման ճշտությամբ գերազանցել է բժիշկներին՝ 88.9% ընդհանուր, իսկ փորձառու բժիշկների հետ ուղիղ համեմատությունում՝ 87.8% ընդդեմ 78.1%-ի, առանց բարձր ծանրության դեղորայքային սխալների:

Ինչն է հավանական, բայց ապացուցված չէ. Որ այս կարողությունը իրական, նախապես չընտրված հիվանդների համար կլինիկական օգուտ կտա, և որ ճշտության տարբերությունը գալիս է ավելի լավ reasoning-ից, ոչ ավելի շատ թեստերից, պատմական քարտին համապատասխանությունից կամ հանրային տվյալների հիշողությունից:

Ինչ չի ցույց տալիս. Որ MIRA-ն աշխատում է ութ ախտորոշումներից դուրս, անվտանգ է իրական ներդրման համար, բժիշկներին գերազանցում է հավասար տեղեկատվությամբ արդար համեմատությունում, փոխարինում է կլինիցիստներին կամ զերծ է ուսուցման տվյալների աղտոտումից:

Հիմնական սահմանափակումները. Հետահայաց, սիմուլացված, միայն տեքստային գնահատում, ութ նախապես ընտրված ախտորոշում, տեղեկատվական անհամաչափություն՝ զգալիորեն ավելի շատ արյան թեստեր, բուժման արդյունքների մի մասի գնահատում պատմական քարտի համաձայնությամբ, և հանրային MIMIC-IV բենչմարք, որը լեզվային մոդելը կարող էր տեսած լինել ուսուցման ընթացքում: Հեղինակներն իրենք սա նշում են որպես կատարողականի հնարավոր վերին սահման:

Որքա՞ն վստահություն պետք է ունենա ընդհանուր ընթերցողը. Բարձր վստահություն, որ MIRA-ն իրական և նոր կարողության ցուցադրում է՝ գործակալ, որը պարզապես չի պատասխանում, այլ *գործում է* բժշկական գրառման միջով: Ցածր վստահություն այն ավելի լայն պնդման հանդեպ, թե այն *ավելի լավ ախտորոշող է, քան բժիշկը*: Այդ պնդումը սահմանափակված է մեկ հետահայաց բենչմարքով և խճճված է թեստերի քանակով, պատմական քարտով գնահատմամբ ու հնարավոր տվյալային աղտոտմամբ: Հեղինակների սեփական եզրակացությունն ամենաճիշտն է՝ իրական հիվանդների խնամքի մասին որևէ բան ասելուց առաջ պետք են

հեռանկարային, իրական միջավայրի ուսումնասիրություններ։

Աղբյուրներ

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