The Clean Paper

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

चिकित्सकीय AI औसतन निजी दिख सकती है, फिर भी कुछ मरीजों को उजागर कर सकती है — खासकर कम प्रतिनिधित्व वाले लोगों को

"गोपनीयता-सुरक्षित" कहे जाने वाले चिकित्सकीय AI मॉडल अक्सर एक औसत संख्या पर टिके होते हैं। यह अध्ययन कहता है कि वह कसौटी पर्याप्त नहीं है। लेखकों ने membership inference हमलों का अध्ययन किया—ऐसे हमले जो यह अनुमान लगाते हैं कि किसी व्यक्ति का रिकॉर्ड मॉडल के प्रशिक्षण डेटा में था या नहीं, और इस तरह किसी बीमारी से जुड़ी संवेदनशील जानकारी का संकेत दे सकते हैं। सात चिकित्सकीय डेटासेट (इमेजिंग, ECG और इलेक्ट्रॉनिक रिकॉर्ड) तथा अनेक मॉडलों में उन्होंने जोखिम को औसत के बजाय मरीज-दर-मरीज मापा। नतीजा औसतन सुरक्षित दिखने वाला मॉडल भी कुछ व्यक्तियों को लगभग पूरी तरह पहचानने दे सकता है (attack AUC ≥ 0.95); सबसे अधिक उजागर लोग व्यवस्थित रूप से कम प्रतिनिधित्व वाले समूहों से आते हैं — अल्पसंख्यक जातीयता, दुर्लभ रोग या असामान्य इमेजिंग — और बड़े मॉडल में समस्या बढ़ती है। लेखक चिकित्सकीय AI छोड़ने को नहीं कहते; वे प्रति-मरीज गोपनीयता मापने, मॉडल तक पहुँच नियंत्रित करने और differential privacy इस्तेमाल करने की सलाह देते हैं। असुविधाजनक निष्कर्ष: "औसतन निजी" होना किसी व्यक्ति की गोपनीयता की गारंटी नहीं है।

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

Paper: *Disparate privacy risks from medical AI*, Moritz Knolle, Georg Kaissis, Daniel Rückert and colleagues (Technical University of Munich; Imperial College London; Hasso Plattner Institute), Nature (2026) · DOI: 10.1038/s41586-026-10688-0

गोपनीयतागारंटीकिसीऔसत से नहीं, हर व्यक्ति से कियागया वादाहै

जब किसीmedical-AI मॉडल को"privacy-preserving" कहाजाताहै, तोयह दावाअक्सर एक हीसंख्यापर टिकाहोताहै: जिन सभीमरीज़ के डेटासे मॉडल train हुआ, उनमें औसतन किसीव्यक्ति कीसदस्यताकाअनुमान लगाने कीसंभावनाकम है। यह आश्वस्त करने वालालगताहै। इस शोधपत्र काशांत लेकिन असुविधाजनक बिंदु है कि यह गलत संख्याहै।

गोपनीयताकोई औसत नहींहै। यह हर व्यक्ति से कियागयावादाहै—कि प्रशिक्षण set में होनाभविष्य में उन्हें उजागर नहीं करेगा। और कोई औसत लगभग सभीके लिए यह वादानिभाते हुए कुछ लोगोंके लिए इसे पूरीतरह तोड़ सकताहै। शोधकर्ताओं ने गोपनीयताजोखिम कोpatient-by-patient, पहले कभीन इस्तेमाल कीगई resolution पर मापा, और ठीक यहीपाया समग्र में सुरक्षित दिखने वाले मॉडल कुछ खास व्यक्तियोंकीसदस्यतालगभग पूरीतरह रिसाव कर सकते हैं—और जिनकीसदस्यता रिसाव होतीहै, वे अनुपात से अधिक वहीलोग हैं जिन्हें पहले से सबसे कम सुरक्षामिलीहुई है।

लेखकोंने क्याकिया

Moritz Knolle के नेतृत्व वालीटीम—Technical विश्वविद्यालय of Munich के Daniel Rückert और Hasso Plattner Institute के Georg Kaissis, साथ में Imperial College London के सहयोगी—ने **सदस्यताअनुमान हमलाना** के प्रसिद्ध गोपनीयताthreat कोलियाऔर सवाल बदल दिया। "डेटासेट में औसतन यह हमलाकितनीबार सफल होताहै?" पूछने कीबजाय उन्होंने पूछा, "*इस खास मरीज़* के लिए एक्सपोज़र कितनाहै?"

उन्होंने यह per-patient विश्लेषण **सात स्थापित चिकित्सकीय डेटासेट** पर चलाया, जिनमें बहुत अलग डेटाtypes थे—चिकित्सकीय इमेजिंग, electrocardiograms और इलेक्ट्रॉनिक स्वास्थ्य रिकॉर्ड—और हर डेटासेट पर 200 मॉडल तक। हर मॉडल में हर मरीज़ के लिए उन्होंने अनुमान लगायाकि attacker कितने विश्वास से बतासकताहै कि उस व्यक्ति का रिकॉर्ड प्रशिक्षण डेटामें थायानहीं, फिर परिणाम कोबीमारीस्थिति, self-reported race, insurance, sex और इमेजिंग प्रोटोकॉल जैसे समूह में बाँटा।

Membership inference attack वास्तव में क्याहै

यहाँरिसाव "मॉडल आपकापूरारिकॉर्ड बाहर उगल देताहै" जैसीचीज़ नहींहै। यह *सदस्यता*के बारे में है—सिर्फ यह तथ्य कि आपकाडेटाप्रशिक्षण set में था।

पहले समझें कि ऐसामॉडल सामान्यत:क्याकरताहै। आप उसे कोई स्कैन—मान लें chest X-ray—देते हैं और वह probabilities लौटाताहै: pneumonia की78% संभावना, साथ में cardiomegaly, oedema, consolidation जैसीreadings। हमलाइसीरोज़मर्राके नैदानिक उत्तर काउपयोग करताहै। कुछ असाधारण नहीं।

कमज़ोरीयह है कि मॉडल अक्सर उन सटीक उदाहरण पर थोड़ाअधिक confident होताहै जिन पर वह train हुआ था, बनिस्बत उन उदाहरण के जिन्हें उसने कभीनहींदेखा। इसे ऐसे student जैसासमझें जिसने चुपके से exam पहले देख लियाहो—practice किए प्रश्न पर उत्तर थोड़ाज़्यादातेज़ीऔर विश्वास से आते हैं। इसीसंकेत से यह पता लगानाकि कोई खास रिकॉर्ड प्रशिक्षण उदाहरण में थायानहीं, "सदस्यताअनुमान" है।

हमलाचरण by चरण ऐसाहै। कोई व्यक्ति जानना चाहताहै कि किसीखास मरीज़ कास्कैन मॉडल प्रशिक्षण में इस्तेमाल हुआ थायानहीं। वह मॉडल से रिकॉर्ड माँगता**नहीं**—उसके पास उम्मीदवार स्कैन पहले से है, मरीज़ काअपना याबहुत करीबीप्रतिलिपि। वह स्कैन कोसामान्य नैदानिक request कीतरह एक बार भेजताहै और उत्तर कावि श्वास टिप्पणीकरताहै। फिर वह देखताहै कि यह विश्वास उस मॉडल जैसालगताहै जिसने स्कैन पर train कियाथायाउस

जैसाजसिने नहींकयिाथा—यह तुलनावह अपने computer पर एक stand-in “संदर्भ मॉडल” train करके सस्ते में सीख सकताहै, बनिावशिष hardware के। यदि वास्तवकि मॉडल इस स्कैन पर असमान्य रूप से sure है—मानोउसे पहचानताहो—तोयह मजबूत साक्ष्य है कि स्कैन प्रशक्षिण डेटामें था।

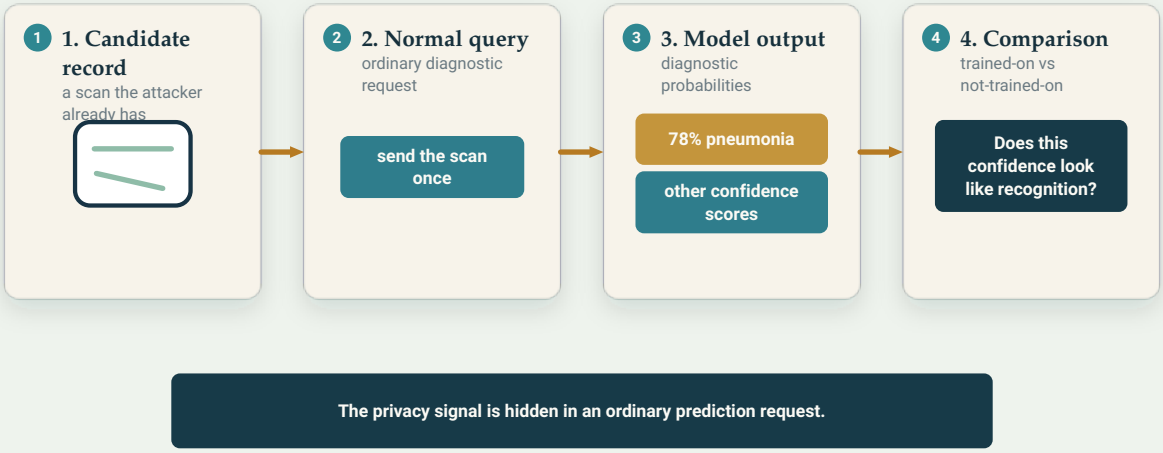
असुवधिाजनक बात यह है कि request हमलाजैसीदखितीहीनहीं यह वहीनैदानकि प्रश्न है जोचकित्सिक कर सकताहै, और गोपनीयतासंकेत समान्य पूर्वानुमान के भीतर छुपाहै। इसीकारण जोखमि ठोस है—हालाँकि अभीतक इसे बतई गई प्रयोगशालामान्यताएँ के तहत दखिायागयाहै, वास्तवकि दुनयिामें पकड़ानहींगया।

यदि रकिॉर्ड खुद रसिद्ध नहींहोतोसदस्यताक्योंमायने रखतीहै? क्योंकि *सदस्यतास्वयं एक तथ्य है*। यदि मॉडल खास cancer immunotherapy पाए मरीज़ पर train हुआ, तोयह confirm करनाकि आपकारकिॉर्ड उसमें है, यह बतासकताहै कि आपकोशायद वहीcancer था—ऐसीबात जसिकाअनुमान insurer याemployer कोकभीनहींलगापाना चहिए। Content बंद रहताहै; बहर नकिलताहै belonging कातथ्य।

लेखक मान्यताएँ साफ़ लखिते हैं—मॉडल कीसमान्य पूर्वानुमान तक पहुँच, उम्मीदवार रकिॉर्ड और attacker काअपना संदर्भ मॉडल—कसिीhow-to के रूप में नहीं, बल्कि तकि threat पर अस्पष्ट डर कीबजाय तर्क कयिाजासके।

A membership attack can hide inside a normal prediction

The attacker does not ask for the record. They already hold a candidate and query the model once.



Mechanism only: no pseudocode, no attack threshold, no recipe.

हमलाकोकसिीवशिष गोपनीयताAPI कीजरूरत नहीं। यदि मॉडल पहले देखे गए रकिॉर्ड पर असमान्य वशि्वास दखिताहै तोसमान्य नैदानकि प्रश्न सदस्यतासंकेत रसिद्ध कर सकतीहै।

उन्हें क्यामलिा

तीन नष्किर्ष, और हर अगलीपहले से अधकि तीखी।

Averages उजगार लोगोकोछपिदेते हैं। समग्र में—यमीआम तरीक़ा—इनमें से कई मॉडल आश्वस्त रूप से private दखिते हैं: अधकिंश मरीज़ के लिए हमलासंयोग से बेहतर नहीकरता। Per-patient view ने अलग कहनीदखिई। Knolle ने अध्ययन announcement में कहाकि पहले के assessments केवल सभीमरीज़ काऔसत जोखमि मापते थे; पहलीबार व्यक्तमिरीज़ के स्तर पर देखने से तस्वीर बहुत अलग नकिली। कुछ लोगोंके लिए हमला**लगभग पूरीतरह सफल** होताहै—लेखक इसे **0.95 याअधकि हमला**AUC के रूप में मापते हैं (0.5 coin toss, 1.0 पूर्ण)—जबकि पूरे डेटासेट काऔसत संयोग से बेहतर नहीं दखिता।

औसत संयोग के पास रह सकताहै, जबकि रकिॉर्ड कीछोटीtail बहुत अधकि उजगार रहतीहै। शोधपत्र काबदिु है कि गोपनीयतामरीज़ level पर जाँच करनीहोगी, केवल समग्र में नहीं।

एक्सपोज़र असमान है—और कम प्रतनिधित्वि वालालोगोंपर ज्यादापड़ताहै। पास-perfect हमलाके प्रतसिबसे vulnerable मरीज़ व्यवस्थति रूप से उन समूह में थे जोडेटामें **कम प्रतनिधित्वि वाला**थे: minority ethnicities, rare-बीमारी phenotypes, unusual इमेजगि characteristics। Electronic-record डेटासेट में Black मरीज़ सबसे vulnerable रकिॉर्ड के बीच अपेक्षति से **31% अधकि बार** आए; mammography set में malignancy के लिए suspicious scans उस danger क्षेत्र में आश्चर्यजनक **+1,179%** overrepresented थे। (ये दोचत्रि उदहरण हैं, पूरीतस्वीर नहीं—शोधपत्र कई समूह में यही skew दखिताहै—और ये छोटीextreme-risk tail के भीतर *समेक्ष* over-नरिूपण हैं: उजगार रकिॉर्ड में ये मरीज़ अपेक्षासे कतिनीअधकि बार आते हैं, न कि Black यासuspicious-mammography मरीज़ काकतिनाहसि्साउजगार है।) मॉडल के पास इन मरीज़ जैसे कम उदहरण होते हैं जनिमें उन्हें "मलिा" सके, इसलिए उनकारकिॉर्ड अलग दखिई देताहै—और अलग दखिनावहीचीज़ है जसिे सदस्यताहमलापकड़ताहै। गोपनीयताfailure यहृच्छकि नहींहै; वह margins पर पहले से मौजूद लोगोंपर केंद्रति होतीहै।

बड़े मॉडल समस्याबढ़ते हैं। पास-perfect हमले के लिए उजगार मरीज़ कीसंख्यामॉडल क्षमताके साथ तेज़ीसे बढ़ी—एक dermatology डेटासेट में हसि्सासबसे छोटे मॉडल में लगभग शून्य से बढ़कर सबसे बड़े में करीब **दस में एक** होगया। चकित्सकीय AI मॉडल जतिने बड़े और capable होते जारहे हैं—पूरे क्षेत्र कीदशिायहीहै—लेखक चेतावनीदेते हैं कि यह खास जोखमि कम नहीं अधकि गंभीर होताहै।

Rückert का सार साफ़ है: यह tolerable जोखिम नहीं है; स्वास्थ्य डेटा अत्यंत sensitive है।

"औसत न सुरक्षित" गलत परीक्षण क्यों है

कम औसत जोखिम के साथ bill of स्वास्थ्य समझने का आकर्षण है। शोधपत्र का मूल सबक है कि यहाँ averaging केवल imprecise नहीं—गलत चीज़ माप रही है।

गोपनीयता गारंटी तभी सार्थक है जब वह सबसे अधिक जोखिम वाले व्यक्ति के लिए भी काम करे, केवल बीच वाले के लिए नहीं। ऐसा मॉडल जिसमें 1,000 में 999 मरीज़ की सदस्यता recover न हो सके लेकिन एक व्यक्ति लगभग पूर्ण तरीके से identify हो सके, उस एक व्यक्ति के लिए किसी अर्थ में "99.9% private" नहीं है। और यदि वह एक व्यक्ति predictably rare-बीमारी मरीज़ या ethnic-minority मरीज़ है, तो मेट्रिक केवल अधूरा नहीं बल्कि चुपचाप discriminatory है। वह बहुमत की सुरक्षा मापता है और उसे सबकी सुरक्षा कह देता है।

यही बदलाव शोधपत्र मजबूर करता है: *यह मॉडल औसत में कितना private है?* से *सबसे उजागर मरीज़ कौन है, और वह किस समूह से है?* तक। ये अलग सवाल हैं, और गोपनीयता के लिए दूसरा ही असली सवाल है।

यह क्या *सिद्ध यादृच्छिक नहीं* करता

- यह **नहीं** कहता कि चिकित्सकीय AI छोड़ देनी चाहिए। लेखक का फ्रेमिंग mitigation है, retreat नहीं। जोखिम मापें और ठीक करें, building बंद न करें।
- इसका अर्थ यह **नहीं** कि हर चिकित्सकीय मॉडल रिसाव कर रहा है या कोई खास तैनात मॉडल हमला हो चुका है। यह दिखाता है कि *जोखिम मौजूद है और असमान रूप से बँटा है*, और मानक समग्र मेट्रिक्स उसे पूरा न कर पाता करते हैं।
- इसका अर्थ यह **नहीं** कि आपके रिकॉर्ड पहले से उजागर हैं। हमला कोखास स्थितियाँ चाहिए—मॉडल पहुँच, उम्मीदवार रिकॉर्ड और attacker की अपनी infrastructure—यह casual क्षमता नहीं है।
- यह मरीज़ *content* रिसाव हेना **नहीं** दिखाता। रिसाव सदस्यता है—शामिल होने का तथ्य—जो अलग वजह से खतरनाक है, न कि रिकॉर्ड बाहर आ जाने के कारण।
- यह disparities को एक सुरक्षी संख्या में **नहीं** समेटता। मजबूत reproducible परिणाम *पैटर्न* है—समग्र मेट्रिक्स व्यक्ति जोखिम को कम दिखाते हैं, और कम प्रतिनिधित्व वाला मरीज़ सबसे ज़्यादा भार उठाते हैं—जो डेटासेट और सैकड़ों मॉडल में दिखा।

साक्ष्य कितना मजबूत है?

यह empirical, पद्धतिगत परिणाम है और मजबूत है। पैटर्न—समग्र गोपनीयता मेट्रिक्स per-patient जोखिम को systematically कम दिखाती हैं, बचा हुआ जोखिम कम प्रतिनिधित्व वाला समूह पर केंद्रित होता है, और मॉडल क्षमता के साथ बढ़ता है—**सात अलग डेटा types वाले डेटासेट** और हर डेटासेट में बड़ी संख्या में मॉडल पर कायम रहा। यही व्यापकता इसे एक-डेटासेट artefact की बजाय credible मापन दावा बनाती है।

दो ईमानदार caveats हैं। पहला, यह *जोखिम* का प्रदर्शन है, लेखक द्वारा बनाए हमले चलाकर मापा गया; यह बताता है कि stated मान्यताएँ के तहत capable attacker *क्या कर सकता है*, न कि वास्तव में कि दुनिया हमले कितनी बार होते हैं। दूसरा, vivid

संख्याएँ—कुछ मरीज़ के लिए पास-perfect success—डिज़ाइन के अनुसार सबसे worst-off व्यक्ति बनते हैं; यही पूरा बिंदु है। उन्हें “tail औसत के संकेत से कहीं भारी है” की तरह पढ़ना चाहिए, “most मरीज़ उजागर हैं” की तरह नहीं।

उचित रुख न चेतावनी है, न dismissal: कई डेटासेट पर किया गया सावधान अध्ययन दिखाता है कि medical-AI गोपनीयता certify करने का मानक तरीका अपने worst मामले को देख ही नहीं पाता—और वही worst मामले उन मरीज़ पर पड़ते हैं जिनके पास पहले से सबसे कम margin है।

यह क्यों मायने रखता है

दो बातें इसे technical footnote से अधिक बनाती हैं।

पहली, यह बदलता है कि “privacy-preserving” कहने की अनुमति किसे मिलनी चाहिए। यदि मॉडल औसत गोपनीयता स्कोर के साथ रिलीज़ किया जाता है, स्कोर सचमुच कम हो सकता है और फिर भी मरीज़ का subset सदस्यता हमले से पास-perfectly identifiable हो सकता है। शोधपत्र की व्यावहारिक demand ठोस है: रिलीज़ से पहले गोपनीयता जोखिम **हर मरीज़ के स्तर पर** assess करें, तैनात मॉडल की पहुँच नियंत्रित करें, और **अंतर गोपनीयता** जैसी techniques इस्तेमाल करें—प्रशिक्षण में थोड़ा, carefully calibrated शोर जोड़ना जो सदस्यता हमले को blunt करता है, मॉडल usefulness की वास्तविक और managed लागत के साथ। गोपनीयता–utility trade-off स्पष्ट है, मुक्त नहीं। “हमने औसत जाँच किया” अब पर्याप्त नहीं होना चाहिए।

दूसरी, यह गोपनीयता को fairness से जोड़ता है। चिकित्सकीय डेटा में जो समूह कम प्रतिनिधित्व वाले हैं—और इसलिए चिकित्सकीय AI से पहले ही खराब serve होते हैं—वही समूह निकलते हैं जिनकी गोपनीयता AI सबसे कम सुरक्षा देती है। क्षेत्र पहली inequity पर कम करते हुए दूसरी को किसी और विभाग का मुद्दा नहीं मान सकता। लोग वही हैं।

Medical-AI गोपनीयता की आश्वस्त कहानी—*हमने माप किया, औसत कम है, सब ठीक है*—यही कहानी शोधपत्र खोलकर अलग करता है। प्रैक्टिस यही किसी को डराकर भगाने के लिए नहीं, बल्कि मानक को वहाँ ले जाने के लिए जहाँ वह होना चाहिए: ऐसा promise जिसे आप तभी दे सकते हैं जब आपने उसे उस व्यक्ति के लिए जाँच किया हो जिसके hurt होने की संभावना सबसे अधिक है।

साफ़ सारांश

सदस्यता अनुमान हमले यह पता लगाने की कोशिश करते हैं कि किसी खास व्यक्ति का रिकॉर्ड AI मॉडल के प्रशिक्षण डेटा में था या नहीं—और क्योंकि सदस्यता खुद revealing हो सकती है (जैसे आपको कोई खास बीमारी थी), यह वास्तव में गोपनीयता रिसाव है भले मूल रिकॉर्ड कभी बाहर न आए। पहले का काम डेटासेट पर ऐसे हमले की औसत सफलता मापता था। इस अध्ययन ने सात चिकित्सकीय डेटासेट—इमेजिंग, ECG, इलेक्ट्रॉनिक रिकॉर्ड—और हर डेटासेट के कई मॉडल में इसे *हर मरीज़ के लिए अलग* मापा और पाया कि समग्र मेट्रिक्स जोखिम को बहुत कम दिखाती हैं: कुछ व्यक्ति लगभग perfectly identify किए जा सकते हैं (हमला AUC ≥ 0.95) भले डेटासेट-wide औसत सुरक्षित लगे; सबसे vulnerable systematically कम प्रतिनिधित्व वाला समूह—minority ethnicity, दुर्लभ बीमारी, unusual इमेजिंग—से हैं; और मॉडल आकार के साथ समस्या बढ़ती है। लेखक चिकित्सकीय AI छोड़ने को नहीं कहते—वे individual-level गोपनीयता मापने, मॉडल पहुँच नियंत्रण और अंतर गोपनीयता की माँग करते हैं। Takeaway: “औसत में private” गोपनीयता गारंटी नहीं है, और जहाँ यह विफल होना होती है वहाँ अक्सर वही मरीज़ होते हैं जिन्हें पहले से सबसे कम सुरक्षा मिली है।

बिना बढ़ा-चढ़ाकर जाँच

पेपर क्या दिखाता है: सात चिकित्सकीय डेटासेट और हर डेटासेट के कई मॉडल में कुछ व्यक्ति के लिए per-patient membership-अनुमान जोखिम समग्र मेट्रिक्स के संकेत से बहुत अधिक है—पास-perfect हमला success (AUC $\geq$ 0.95) तक—और बचाहुआ जोखिम कम प्रतिनिधित्व वाले समूह पर disproportionate रूप से पड़ता है तथा मॉडल क्षमता के साथ बढ़ता है।

क्या संभव है लेकिन सिद्ध नहीं कि वास्तविक दुनिया attackers तैनात क्लिनिकल मॉडल के विरुद्ध इसे पहले से इस्तेमाल कर रहे हैं; अध्ययन stated मान्यताएँ में *क्षमता और उसका distribution* दिखाती है, वास्तविक दुनिया में हमले की frequency नहीं।

यह क्या नहीं दिखाता कि चिकित्सकीय AI छोड़ देनी चाहिए; कि हर मॉडल रिसाव करता है या कोई विशिष्ट तैनात मॉडल breach हो चुका है; कि patient-record *contents* सदस्यता की बजाय उजागर करना होते हैं; या disparity की एक ही quantified संख्या—मजबूत परिणाम पैटर्न है, कोई एक संख्या नहीं।

मुख्य सीमाएँ: लेखकों के अपने हमले से worst-case जोखिम माना गया, देखा गया incidents नहीं; dramatic चित्र डिज़ाइन के अनुसार सबसे उजागर व्यक्ति बताते हैं; और सटीक समूह disparities को एक संख्या में समेटने की बजाय डेटासेट में संगत पैटर्न के रूप में दिखाया गया है।

सामान्य पाठक को कितना भरोसा रखना चाहिए? इस पर उच्च भरोसा कि समग्र गोपनीयता मेट्रिक्स व्यक्ति जोखिम को कम दिखाती हैं, बचाहुआ जोखिम कम प्रतिनिधित्व वाला मरीज़ पर केंद्रित होता है और मॉडल आकार के साथ बढ़ता है—यह कई डेटासेट और मॉडल में दिखाया गया है। ऐसे हमले वास्तविक दुनिया में कितनी बार होते हैं, इस पर मध्यम भरोसा क्योंकि अध्ययन इसे माप्त नहीं। सुरक्षित पढ़ाई: न "चिकित्सकीय AI आपका डेटा रिसाव करती है," न "गोपनीयता हल करना हो गई"; बल्कि "माप्त गोपनीयता जाँच अपने worst मामले को नहीं देखती, और वे मामले सबसे vulnerable मरीज़ पर पड़ते हैं—इसलिए जाँच बदलनी होगी।"

स्रोत

Knolle et al., Disparate privacy risks from medical AI, Nature (2026)
https://doi.org/10.1038/s41586-026-10688-0

Technical University of Munich — press release, 'Study reveals privacy risks in medical AI' (2026)
https://www.tum.de/en/news-and-events/all-news/press-releases/details/study-reveals-privacy-risks-in

Medical Xpress (2026) — coverage of the study
https://medicalxpress.com/news/2026-06-patient-groups-vulnerable-privacy-medical.html