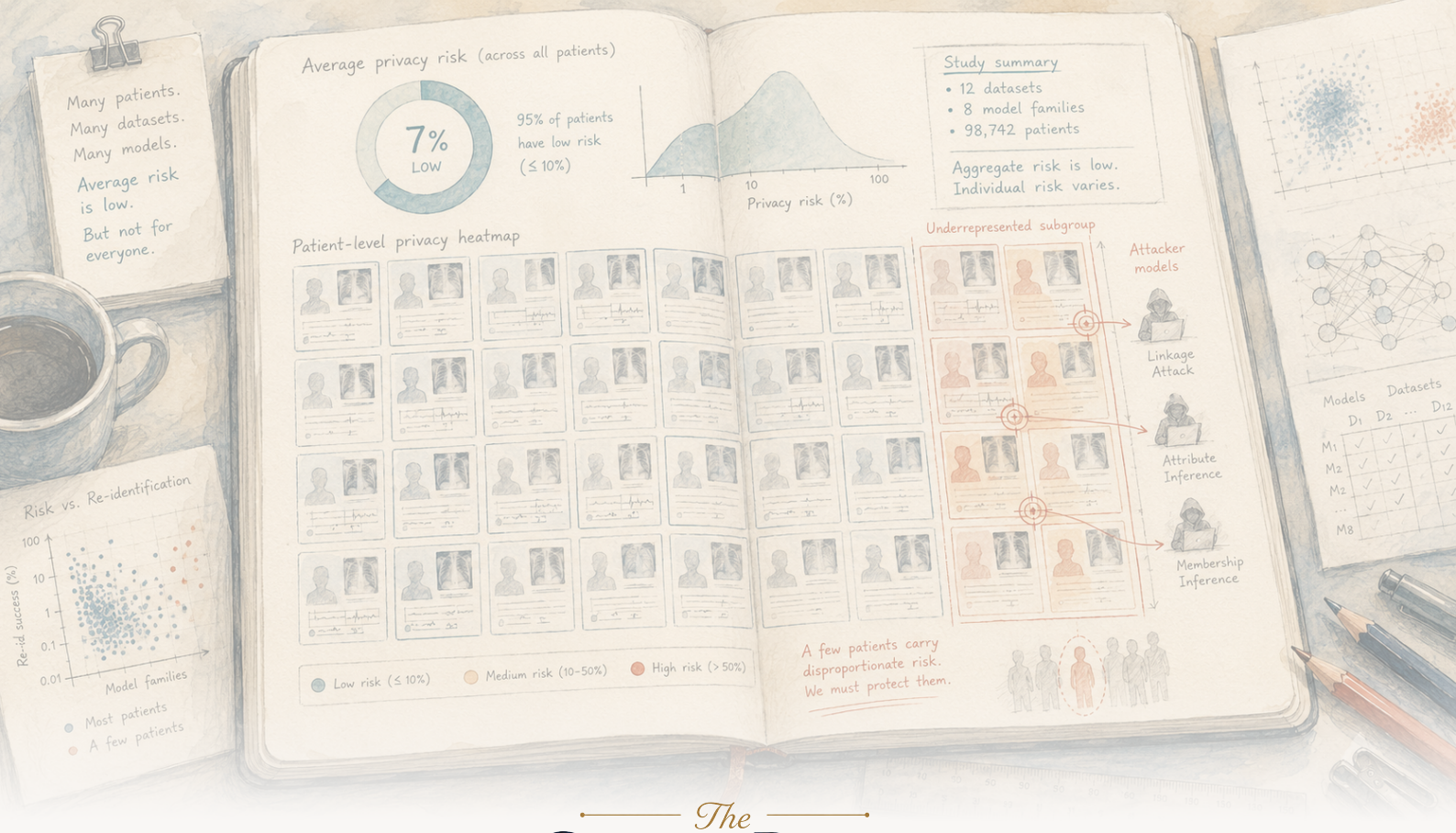


The
CLEAN PAPER

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

پرائیویسی ضمانت اوسط سے نہیں، ہر فرد سے کیا گیا وعدہ ہے

رکنیت استنباط حملے یہ جاننے کی کوشش کرتے ہیں کہ کسی مخصوص *person's* ریکارڈ اے آئی ماڈل کے تربیت ڈیٹا میں تھا یا نہیں۔ چونکہ رکنیت خود *fact sensitive* ہو سکتی ہے — مثلاً کسی بیماری یا علاج سے تعلق ظاہر کر سکتی ہے — یہ *genuine* پرائیویسی رساؤ ہے، چاہے *underlying* ریکارڈ *content* باہر نہ آئے۔ *Previous* کام نے حملہ *success* کو ڈیٹا سیٹ کے اوسط پر پیمائش کیا۔ اس مطالعہ نے اسے *per** مریض** پیمائش کیا، سات طبی ڈیٹا سیٹس (امیجنگ، *ECG*، الیکٹرانک ریکارڈز) اور بہت سے ماڈلز میں، اور پایا کہ مجموعی *metrics* خطرہ کو بہت کم دکھا سکتے ہیں: کچھ *individuals* تقریباً *perfectly* شناخت کرنا ہو سکتے ہیں (حملہ *AUC* ≥ 0.95) حالانکہ ڈیٹا سیٹ-*wide* اوسط محفوظ لگے؛ سب سے زیادہ *systematically vulnerable* کم نمائندگی والے گروپس سے آتے ہیں؛ اور ماڈل حجم کے ساتھ مسئلہ بڑھتی ہے۔ مصنفین طبی اے آئی چھوڑنے کی بات نہیں کرتے — وہ فرد-سطح پرائیویسی پیمائش، ماڈل-رسائی *controls* اور *differential* پرائیویسی کی سفارش کرتے ہیں۔ *Takeaway*: "اوسط پر *private*" پرائیویسی ضمانت نہیں، اور ناکامی زیادہ تر انہی لوگوں پر آتی ہے جو پہلے ہی سب سے کم *protected* ہیں۔

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: 0-10688-026-s41586/1038.10

پرائیویسی کی ضمانت اوسط کے لیے نہیں، ہر فرد سے کیا گیا وعدہ ہے

جب کسی طبی اے آئی ماڈل کو "پرائیویسی محفوظ رکھنے والا" کہا جاتا ہے تو یہ دعویٰ عموماً ایک مجموعی عدد پر قائم ہوتا ہے: جن مریضوں کے ڈیٹا سے ماڈل کو تربیت دی گئی، ان سب کو ملا کر کسی ایک فرد کی رکنیت پہچانے جانے کا اوسط امکان کم ہے۔ یہ تسلی بخش لگتا ہے۔ اس مقالے کا خاموش مگر بے آرام کرنے والا نکتہ یہ ہے کہ پرائیویسی کے لیے یہی غلط عدد ہے۔

پرائیویسی کوئی اوسط نہیں؛ یہ ہر فرد سے الگ وعدہ ہے کہ تربیتی مجموعے میں شامل ہونا بعد میں اسی کے خلاف استعمال نہیں ہوگا۔ اوسط تقریباً سب کے لیے یہ وعدہ پورا کرتے ہوئے چند لوگوں کے لیے اسے مکمل طور پر توڑ سکتا ہے۔ محققین نے پہلی بار خطرے کو مریض بہ مریض اتنی باریک سطح پر ناپا، اور یہی پایا: مجموعی طور پر محفوظ نظر آنے والے ماڈلز بعض مخصوص افراد کی رکنیت تقریباً کامل درستگی سے ظاہر کر سکتے ہیں—اور غیر متناسب طور پر وہی لوگ زیادہ متاثر ہوتے ہیں جو پہلے ہی کم محفوظ ہیں۔

مصنفین نے کیا کیا

Knolle Moritz کی قیادت میں، Rückert Daniel (دونوں تکنیکی University of، Plattner (Hasso Kaissis Georg Munich) اور Imperial College London) کے ساتھیوں پر مشتمل ٹیم نے پرائیویسی کے ایک معروف خطرے، **رکنیت کے استنباط کے حملے**، کو لیا اور سوال بدل دیا۔ "پورے ڈیٹا سیٹ میں اوسطاً یہ حملہ کتنی بار کامیاب ہوتا ہے؟" کے بجائے انہوں نے پوچھا: "**اس خاص مریض** کے لیے خطرہ کتنا ہے؟"

انہوں نے یہ فی مریض تجزیہ **سات معروف طبی ڈیٹا سیٹس** پر کیا، جن میں مختلف قسم کا ڈیٹا شامل تھا—طبی امیجنگ، electrocardiograms اور الیکٹرانک صحت ریکارڈز—اور ہر ڈیٹا سیٹ کے لیے 200 تک ماڈلز کا جائزہ لیا۔ ہر مریض اور ہر ماڈل کے لیے انہوں نے تخمینہ لگایا کہ حملہ آور کتنے اعتماد سے بتا سکتا ہے کہ اس شخص کا ریکارڈ تربیتی ڈیٹا میں شامل تھا یا نہیں، پھر نتائج کو بیماری کی حالت، خود بتائی گئی نسل، بیمہ، جنس اور امیجنگ طریقۂ کار جیسے گروہوں کے لحاظ سے تقسیم کیا۔

رکنیت کے استنباط کا حملہ دراصل کیا ہے؟

ہے—صرف **رکنیت** چیز والی رسنے یہاں ہے۔" دیتا نکال باھر ریکارڈ طبی پورا کا آپ "ماڈل کہ ہوتا نہیں طرح اس رساؤ یہاں تھا۔ شامل میں مجموعے تربیتی ڈیٹا کا آپ کہ حقیقت یہ

پہلے یہ سمجھیں کہ ایسا ماڈل معمول میں کیا کرتا ہے۔ آپ اسے مثلاً سینے کا ایکسرے دیتے ہیں اور وہ امکانات واپس دیتا ہے: نمونیا کا 78 فیصد امکان، ساتھ cardiomegaly، consolidation oedema وغیرہ کے اسکور۔ حملہ اسی روزمرہ تشخیصی جواب کو استعمال کرتا ہے؛ کسی خاص یا خفیہ رسائی کی ضرورت نہیں۔

کمزوری یہ ہے کہ ماڈل عموماً ان عین مثالوں پر کچھ **زیادہ پراعتماد** ہوتا ہے جن پر اسے تربیت دی گئی ہو، بہ نسبت ان مثالوں کے جو اس نے پہلے نہیں دیکھیں۔ اسے ایسے طالب علم سے تشبیہ دی جا سکتی ہے جس نے امتحان کے کچھ سوال پہلے ہی دیکھ لیے ہوں: ان سوالوں کے جواب کچھ زیادہ تیزی اور اعتماد سے آتے ہیں۔ اسی اضافی اعتماد سے اندازہ لگانا کہ کوئی خاص ریکارڈ ماڈل کی تربیتی مثالوں میں شامل تھا یا نہیں، "رکنیت کا استنباط" ہے۔

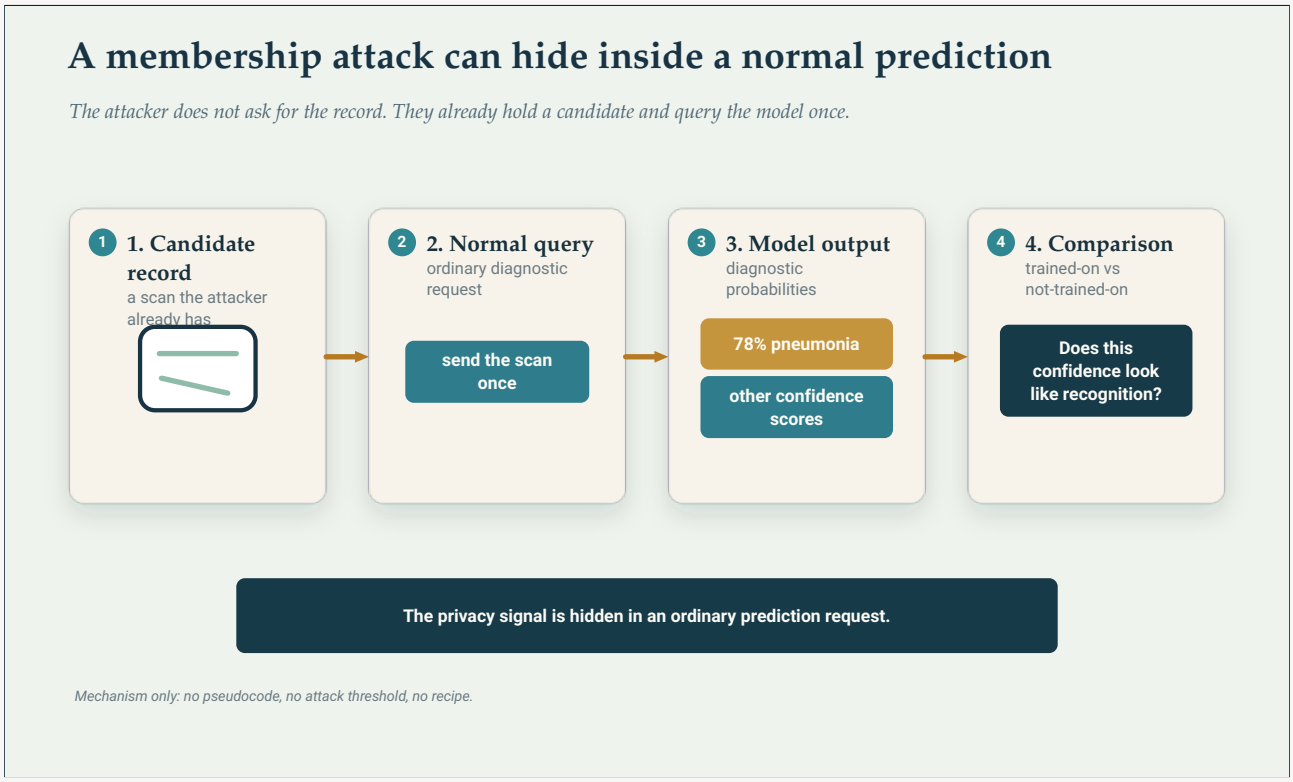
حملہ مرحلہ وار یوں ہوتا ہے۔ کسی شخص کے پاس ایک امکانی اسکین موجود ہے اور وہ جاننا چاہتا ہے کہ آیا یہی اسکین ماڈل کی تربیت میں استعمال ہوا تھا۔ وہ ماڈل سے ریکارڈ نہیں مانگتا؛ اسکین اس کے پاس پہلے سے ہوتا ہے، مثلاً مریض کا اپنا اسکین یا اس کی بہت قریبی نقل۔ وہ اسے عام تشخیصی سوال کے طور پر ماڈل کو دیتا ہے اور جواب کے اعتماد کو دیکھتا ہے۔ پھر اس اعتماد کا موازنہ اس نمونے سے کرتا ہے جو ایسے ریکارڈز پر بنتا ہے جنہیں ماڈل نے تربیت میں دیکھا

ہو اور ایسے ریکارڈز پر جنہیں نہ دیکھا ہو۔ یہ فرق سیکھنے کے لیے حملہ آور اپنا ایک حوالہ ماڈل عام کمپیوٹر پر تربیت دے سکتا ہے۔ اگر اصل ماڈل اس اسکین پر غیر معمولی یقین دکھائے—گویا اسے پہچانتا ہو—تو یہ اس کے تربیتی ڈیٹا میں شامل ہونے کا مضبوط اشارہ بن سکتا ہے۔

پریشان کن بات یہ ہے کہ درخواست ظاہری طور پر حملہ لگتی ہی نہیں؛ یہ وہی تشخیصی سوال ہے جو ایک معالج بھی پوچھ سکتا ہے، اور پرائیویسی کا اشارہ ایک عام پیش گوئی کے اندر چھپا ہوتا ہے۔ اسی لیے خطرہ عملی نوعیت کا ہے، اگرچہ اب تک اسے واضح تجربہ گاہی مفروضوں کے تحت دکھایا گیا ہے، حقیقی دنیا میں پکڑے گئے حملوں کی شرح کے طور پر نہیں۔

اگر اصل ریکارڈ باہر نہیں آتا تو رکنیت اہم کیوں ہے؟ کیونکہ **رکنیت خود ایک حقیقت ہے**۔ فرض کریں ماڈل صرف ایسے مریضوں کے ڈیٹا پر تربیت یافتہ ہو جنہیں کسی خاص سرطان کی immunotherapy ملی تھی۔ اگر کسی شخص کی رکنیت ثابت ہو جائے تو اس سے اس سرطان یا علاج کے بارے میں حساس نتیجہ اخذ کیا جا سکتا ہے—ایسی بات جو کسی بیمہ کمپنی یا آجر کو معلوم نہیں ہونی چاہیے۔ ریکارڈ کا متن بند رہتا ہے؛ تربیتی مجموعے سے تعلق کی حقیقت باہر آتی ہے۔

مصنفین اپنے مفروضے صاف بیان کرتے ہیں—ماڈل کی معمول کی پیش گوئیوں تک رسائی، ایک امکانی ریکارڈ، اور حملہ آور کا اپنا حوالہ ماڈل—تاکہ خطرے پر واضح انداز میں بحث ہو سکے، نہ کہ مبہم خوف کی بنیاد پر۔



حملے کو کسی خاص پرائیویسی اے پی آئی کی ضرورت نہیں۔ اگر ماڈل پہلے دیکھے گئے ریکارڈ پر غیر معمولی اعتماد دکھاتا ہے تو ایک معمول کا تشخیصی سوال بھی رکنیت کا اشارہ ظاہر کر سکتا ہے۔

Original Aurora diagram — The Clean Paper CC BY 0.4

کیا ملا

تین نتائج سامنے آئے، اور ہر اگلا نتیجہ پہلے سے زیادہ اہم تھا۔

اوسط سب سے زیادہ متاثر لوگوں کو چھپا دیتا ہے۔ مجموعی طریقے سے، جو عام پیمانہ ہے، بہت سے ماڈلز تسلی بخش حد تک محفوظ دکھائی دیتے ہیں: زیادہ تر مریضوں کے لیے حملہ محض اتفاق سے بہتر کارکردگی نہیں دکھاتا۔ لیکن فی مریض منظر مختلف تھا۔ Knolle نے مطالعے کے اعلان میں کہا کہ پچھلی جانچیں صرف تمام مریضوں کا اوسط خطرہ ناپتی تھیں، جبکہ یہاں پہلی بار انفرادی مریض کی سطح پر خطرہ دیکھا گیا۔ بعض افراد کے لیے حملہ **تقریباً کامل** کامیابی دکھاتا ہے—مصنفین نے اسے **95.0 یا اس سے زیادہ** AUC کے طور پر ناپا (5.0 سکہ اچھالنے کے برابر، 0.1 کامل)—حالانکہ پورے ڈیٹا سیٹ کا اوسط محض اتفاق کے قریب تھا۔



اوسط اتفاق کے قریب ہو سکتا ہے، جبکہ ریکارڈز کا ایک چھوٹا حصہ بہت زیادہ سامنے ہو۔ مقالے کا بنیادی نکتہ یہی ہے کہ پرائیویسی صرف مجموعی طور پر نہیں، مریض کی سطح پر بھی جانچنی چاہیے۔

Original Aurora diagram — The Clean Paper CC BY 0.4

خطرہ برابر تقسیم نہیں ہوتا—اور زیادہ تر کم نمائندگی والے گروہوں پر پڑتا ہے۔ تقریباً کامل حملے کے لیے سب سے زیادہ کمزور مریض منظم طور پر انہی گروہوں سے تھے جو ڈیٹا میں **کم نمائندگی** رکھتے تھے: اقلیتی نسلی گروہ، نایاب بیماریوں کی مخصوص صورتیں، یا غیر معمولی امیجنگ خصوصیات۔ الیکٹرانک ریکارڈ والے ڈیٹا سیٹ میں سیاہ مریض سب سے زیادہ خطرے والے ریکارڈز میں توقع سے **31 فیصد زیادہ** تھے؛ mammography ڈیٹا سیٹ میں malignancy کے لیے مشکوک اسکینز انتہائی خطرے والے حصے میں **1,179 فیصد زیادہ** نمائندگی رکھتے تھے۔ یہ دو مثالیں ہیں، پورا نتیجہ نہیں، اور یہ نسبتیں اس چھوٹے انتہائی خطرے والے حصے کے اندر اضافی نمائندگی بیان کرتی ہیں—یہ نہیں کہ سیاہ یا مشکوک mammography والے اتنے فیصد مریض لازماً متاثر ہیں۔ کم نمائندگی والے مریضوں کے لیے مشابہ مثالیں کم ہوتی ہیں، اس لیے ان کے ریکارڈ زیادہ نمایاں رہ سکتے ہیں، اور یہی نمایاں ہونا رکنیت کے حملے کے لیے مفید اشارہ بنتا ہے۔

بڑے ماڈلز میں خطرہ بڑھا۔ تقریباً کامل حملوں کے سامنے آنے والے مریضوں کی تعداد ماڈل کی صلاحیت بڑھنے کے ساتھ تیزی سے بڑھی۔ ایک dermatology ڈیٹا سیٹ میں یہ حصہ سب سے چھوٹے ماڈل میں تقریباً صفر تھا، مگر سب سے بڑے ماڈل میں تقریباً **ہر دس میں سے ایک** مریض تک پہنچ گیا۔ مصنفین کے مطابق طبی اے آئی جس سمت بڑھ رہی ہے—بڑے اور زیادہ قابل ماڈلز—وہ اس مخصوص خطرے کو کم نہیں بلکہ بڑھا سکتی ہے۔

Rückert کا خلاصہ سخت ہے: “یہ قابلِ قبول خطرہ نہیں۔ صحت کا ڈیٹا نہایت حساس ہوتا ہے۔”

“اوسطاً محفوظ” ہونا غلط جانچ کیوں ہے

کم اوسط خطرہ دیکھ کر یہ سمجھنا آسان ہے کہ مسئلہ حل ہو گیا۔ مقالے کا بنیادی سبق یہ ہے کہ یہاں اوسط لینا صرف غیر دقیق نہیں؛ یہ **غلط چیز ناپتا ہے۔**

پرائیویسی کی ضمانت تب معنی رکھتی ہے جب وہ سب سے زیادہ خطرے والے فرد کے لیے بھی قائم رہے۔ اگر 1,000 میں 999 مریض شناخت نہ کیے جا سکیں مگر ایک مریض تقریباً کامل درستگی سے پہچانا جا سکے تو اس ایک شخص کے لیے ماڈل “99.9 فیصد محفوظ” نہیں۔ اور اگر وہ فرد منظم طور پر نایاب بیماری والا یا کسی نسلی اقلیت سے تعلق رکھنے والا مریض ہو تو پیمانہ صرف نامکمل نہیں، خاموشی سے امتیازی بھی بن جاتا ہے: اکثریت کی حفاظت ناپ کر اسے سب کی حفاظت کا نام دے دیتا ہے۔

مقالہ سوال ہی بدل دیتا ہے: ماڈل اوسطاً کتنا محفوظ ہے؟ کے بجائے سب سے زیادہ سامنے آنے والا مریض کون ہے، اور وہ کس گروہ سے تعلق رکھتا ہے؟۔ یہ دو مختلف سوال ہیں، اور فرد کی پرائیویسی کے لیے دوسرا سوال زیادہ اہم ہے۔

یہ کیا ثابت یا دعویٰ نہیں کرتا

- یہ **نہیں** کہتا کہ طبی اے آئی چھوڑ دینی چاہیے۔ مصنفین کا مقصد خطرہ کم کرنا ہے، پیچھے ہٹنا نہیں: خطرہ ناپیں اور اسے کم کریں، ٹیکنالوجی بنانا بند نہ کریں۔
- اس کا مطلب یہ **نہیں** کہ ہر طبی ماڈل معلومات رسا رہا ہے یا کسی خاص تعینات ماڈل پر حملہ ہو چکا ہے۔ مطالعہ دکھاتا ہے کہ خطرہ موجود ہے اور برابر تقسیم نہیں، اور عام مجموعی پیمانے اسے چھپا سکتے ہیں۔
- اس کا مطلب یہ **نہیں** کہ آپ کے ریکارڈ پہلے ہی سامنے آ چکے ہیں۔ حملے کے لیے خاص شرائط درکار ہیں—ماڈل تک رسائی، امکانی ریکارڈ اور حملہ آور کا اپنا بنیادی ڈھانچا—یہ کوئی اتفاقی صلاحیت نہیں۔
- یہ مریض کے ریکارڈ کا **مواد** رسنے کو نہیں دکھاتا۔ رسنے والی چیز رکنیت ہے—شمولیت کی حقیقت—جو ایک مختلف وجہ سے حساس ہو سکتی ہے۔
- یہ تفاوت کو کسی ایک سرخی والے عدد تک **محدود** نہیں کرتا۔ مضبوط اور بار بار نظر آنے والا نتیجہ ایک نمونہ ہے: مجموعی پیمانے فرد کے خطرے کو کم ظاہر کرتے ہیں، اور کم نمائندگی والے مریض زیادہ بوجھ اٹھاتے ہیں—یہ پیٹرن کئی ڈیٹا سیٹس اور سینکڑوں ماڈلز میں نظر آیا۔

شواہد کتنے مضبوط ہیں؟

یہ ایک تجرباتی اور طریقۂ کار سے متعلق نتیجہ ہے، اور کافی مضبوط ہے۔ تینوں نمونے—مجموعی پرائیویسی پیمانوں کا فی مریض خطرہ کم دکھانا، باقی رہ جانے والے خطرے کا کم نمائندگی والے گروہوں میں مرتکز ہونا، اور ماڈل کی صلاحیت کے ساتھ اس کا

بڑھنا—**سات مختلف اقسام کے طبی ڈیٹا سیٹس** اور ہر ڈیٹا سیٹ کے بہت سے ماڈلز میں برقرار رہے۔ یہی وسعت اسے ایک ہی ڈیٹا سیٹ کی اتفاقی خصوصیت کے بجائے قابلِ اعتماد پیمائشی نتیجہ بناتی ہے۔

دو اہم احتیاطیں ہیں۔ پہلی، یہ خطرے کی صلاحیت کا مظاہرہ ہے: مصنفین نے اپنے بنائے ہوئے حملے چلا کر ناپا کہ بیان کردہ مفروضوں کے تحت ایک قابل حملہ آور کتنی اچھی کارکردگی دکھا سکتا ہے؛ اس سے حقیقی دنیا میں حملوں کی تعدد معلوم نہیں ہوتی۔ دوسری، بعض مریضوں کے لیے تقریباً کامل کامیابی جیسے نمایاں اعداد جان بوجھ کر **سب سے زیادہ متاثر افراد** کو بیان کرتے ہیں—یہی مطالعے کا نکتہ ہے۔ انہیں "زیادہ تر مریض سامنے آ گئے" کے طور پر نہیں بلکہ "انتہائی خطرے والا حصہ اوسط سے کہیں زیادہ بھاری ہے" کے طور پر پڑھنا چاہیے۔

مناسب مؤقف نہ خوف پھیلانا ہے نہ مسئلہ رد کرنا: یہ کئی ڈیٹا سیٹس پر کیا گیا محتاط مطالعہ ہے جو دکھاتا ہے کہ طبی اے آئی کی پرائیویسی جانچنے کا عام مجموعی طریقہ اپنے ہی بدترین حالات کو نہیں دیکھ پاتا—اور وہ بدترین حالات انہی مریضوں میں زیادہ آتے ہیں جن کے پاس پہلے ہی سب سے کم حفاظتی گنجائش ہے۔

یہ کیوں اہم ہے

دو باتیں اسے محض تکنیکی حاشیے سے بڑا مسئلہ بناتی ہیں۔

پہلی، یہ بدلتا ہے کہ "پرائیویسی محفوظ رکھنے والا" کہنے کے لیے کیا کافی سمجھا جائے۔ کسی ماڈل کا اوسط پرائیویسی اسکور واقعی کم ہو سکتا ہے، پھر بھی رکنیت کے حملے کے سامنے مریضوں کا ایک چھوٹا گروہ تقریباً کامل طور پر شناخت ہو سکتا ہے۔ مقالے کا عملی مطالبہ واضح ہے: اجرا سے پہلے پرائیویسی کا خطرہ **ہر مریض کی سطح پر** جانچیں، تعینات ماڈلز تک رسائی قابو میں رکھیں، اور **privacy dffierential** جیسے طریقے استعمال کریں—یعنی تربیت کے دوران ناپ تول کر تھوڑا شور شامل کرنا، جو رکنیت کے حملے کو کمزور کرتا ہے مگر ماڈل کی افادیت پر حقیقی اور قابلِ انتظام قیمت بھی رکھتا ہے۔ "ہم نے اوسط دیکھ لیا" کو مکمل جانچ نہیں سمجھنا چاہیے۔

دوسری، یہ پرائیویسی کو انصاف کے مسئلے سے جوڑتا ہے۔ طبی ڈیٹا میں جو گروہ کم نمائندگی رکھتے ہیں—اور جنہیں طبی اے آئی پہلے ہی کم اچھی خدمت دے سکتی ہے—وہی گروہ یہاں کم محفوظ بھی نکلتے ہیں۔ اگر شعبہ پہلی نابرابری پر کام کر رہا ہے تو دوسری کو کسی اور کا مسئلہ نہیں کہہ سکتا۔ دونوں کے مرکز میں وہی لوگ ہیں۔

طبی اے آئی کی پرائیویسی کی تسلی بخش کہانی یہ تھی: ہم نے ناپا، اوسط کم ہے، سب ٹھیک ہے۔ یہ مقالہ اسی کہانی کو کھولتا ہے—ٹیکنالوجی سے لوگوں کو دور کرنے کے لیے نہیں، بلکہ معیار کو وہاں لے جانے کے لیے جہاں اسے ہونا چاہیے: ایسا وعدہ جسے پورا کہنے سے پہلے سب سے زیادہ متاثر ہونے والے فرد کے لیے جانچنا ضروری ہو۔

سادہ خلاصہ

رکنیت کے استنباط کے حملے یہ جاننے کی کوشش کرتے ہیں کہ کسی خاص شخص کا ریکارڈ اے آئی ماڈل کے تربیتی ڈیٹا میں تھا یا نہیں۔ چونکہ رکنیت خود حساس بات ظاہر کر سکتی ہے—مثلاً یہ کہ کسی شخص کو خاص بیماری تھی—اس لیے یہ حقیقی پرائیویسی رساؤ ہے، چاہے اصل ریکارڈ کبھی باہر نہ آئے۔ پچھلا کام ایسے حملوں کی کامیابی کو پورے ڈیٹا سیٹ پر **اوسطاً** ناپتا تھا۔ اس مطالعے نے سات طبی ڈیٹا سیٹس—امیجنگ، ECG اور الیکٹرانک ریکارڈز—اور ہر ایک میں کئی ماڈلز پر خطرہ **فی مریض** ناپا۔ نتیجہ یہ تھا کہ مجموعی پیمانے خطرے کو بہت کم ظاہر کر سکتے ہیں: کچھ افراد تقریباً کامل طور پر پہچانے جا سکتے ہیں (حملے کا

AUC ≥ 95.0)، سب سے کمزور افراد منظم طور پر کم نمائندگی والے گروهوں سے هوتے هیں، اور ماڈل کے سائز کے ساتھ مسئلہ بڑھتا ھے۔ مصنفین طبی اے آئی چھوڑنے کی بات نہیں کرتے؛ وہ فرد کی سطح پر پرائیویسی ناپنے، ماڈل تک رسائی قابو کرنے اور privacy dffierential استعمال کرنے کی تجویز دیتے هیں۔ اصل سبق: "اوسطاً محفوظ" هونا پرائیویسی کی ضمانت نہیں۔

بلا مبالغہ جائزہ

مقالے کیا دکھاتا ھے: سات طبی ڈیٹا سیٹس اور بہت سے ماڈلز میں بعض افراد کے لیے فی مریض رکنیت کے استنباط کا خطرہ مجموعی پیمانوں کے اندازے سے کہیں زیادہ ھے—کچھ کے لیے حملہ تقریباً کامل کامیابی (AUC ≥ 95.0) تک پہنچتا ھے—اور باقی رہ جانے والا خطرہ غیر متناسب طور پر کم نمائندگی والے گروهوں پر پڑتا اور ماڈل کی صلاحیت کے ساتھ بڑھتا ھے۔

کیا قرینِ قیاس ھے مگر ثابت نہیں: حقیقی دنیا کے حملہ آور تعینات طبی ماڈلز کے خلاف پہلے هی اسے استعمال کر رهے هیں۔ مطالعہ بیان کردہ مفروضوں کے تحت صلاحیت اور اس کی تقسیم دکھاتا ھے، حقیقی حملوں کی تعداد نہیں۔

کیا نہیں دکھاتا: یہ نہیں کہ طبی اے آئی چھوڑ دینی چاھیے؛ نہ یہ کہ هر ماڈل رساؤ کرتا ھے یا کوئی خاص تعینات ماڈل توڑا جا چکا ھے؛ نہ یہ کہ مریض کے ریکارڈ کا **مواد** باھر آتا ھے؛ اور نہ تفاوت کو کسی ایک عدد میں سمیٹا گیا ھے—مضبوط نتیجہ مجموعی نمونہ ھے۔

اهم حدود: بدترین صورت کا خطرہ مصنفین کے اپنے حملوں سے ناپا گیا، حقیقی واقعات سے نہیں؛ نمایاں اعداد جان بوجھ کر سب سے زیادہ متاثر افراد کو بیان کرتے هیں؛ اور مختلف گروهوں کے عین تفاوت کو ایک واحد عدد میں سمیٹنے کے بجائے کئی ڈیٹا سیٹس میں مستقل نمونے کے طور پر دکھایا گیا ھے۔

عام قاری کو کتنا اعتماد هونا چاھیے؟ بلند اعتماد کہ مجموعی پرائیویسی پیمانے فرد کے خطرے کو کم ظاهر کر سکتے هیں، باقی خطرہ کم نمائندگی والے مریضوں میں زیادہ مرتکز هوتا ھے، اور ماڈل کے سائز کے ساتھ بڑھتا ھے—یہ کئی ڈیٹا سیٹس اور ماڈلز میں دکھایا گیا ھے۔ حقیقی دنیا میں ایسے حملوں کی تعداد پر درمیانہ اعتماد، کیونکہ یہ مطالعہ اسے نہیں ناپتا۔ محفوظ تعبیر یہ ھے: نہ "طبی اے آئی آپ کا ڈیٹا لازماً رسا دیتی ھے"، نہ "پرائیویسی حل هو چکی ھے"، بلکہ "عام پرائیویسی جانچ اپنے هی بدترین حالات کو نہیں دیکھتی، اس لیے جانچ کا طریقہ بدلنا چاھیے۔"

ماخذ

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

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

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