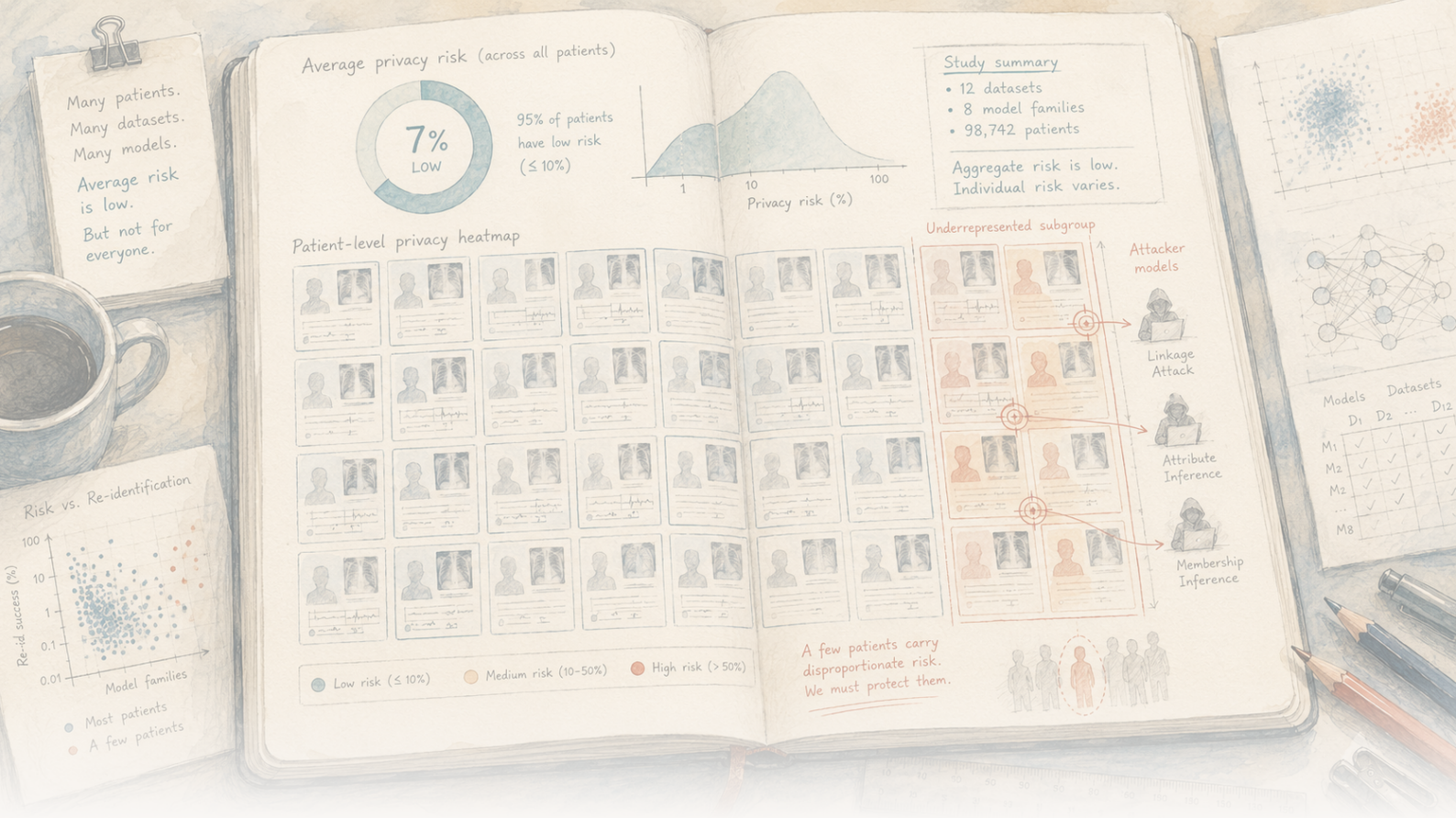


The Clean Paper

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

Medical AI อาจดูเป็นส่วนตัวโดยเฉลี่ย แต่ยังเปิดเผยผู้ป่วยบางคน — โดยเฉพาะกลุ่มที่มีตัวแทนน้อย

คำว่า medical-AI 'privacy-preserving' มักตั้งอยู่บนค่าเฉลี่ยหนึ่งตัว งานนี้เสนอว่านั่นเป็นการทดสอบผิดแบบ ในการศึกษา membership inference attack — ซึ่งเผยว่า record ของบุคคลหนึ่งอยู่ในข้อมูลฝึกหรือไม่ และจึงอาจบอกใบ้ว่าเขามีโรคอะไร — ผู้เขียนวัดความเสี่ยงรายผู้ป่วยแทน aggregate ใน medical dataset 7 ชุด (imaging, ECG, electronic record) และหลายโมเดลต่อชุด รูปแบบที่พบคือ โมเดลที่ดูปลอดภัยโดยเฉลี่ยยังอาจให้ผู้โจมตีระบุบุคคลบางคนได้เกือบสมบูรณ์ (attack AUC ≥ 0.95); คนที่เปิดเผยเป็นกลุ่ม underrepresented อย่างเป็นระบบ (ชนกลุ่มน้อย โรคหายาก imaging ผิดปกติ); และแย่ลงเมื่อโมเดลโต ผู้เขียนไม่ได้บอกให้เลิก medical AI แต่ให้วัด privacy ต่อคน ควบคุม model access และใช้ differential privacy แก่นที่ไม่สบายใจ: 'private โดยเฉลี่ย' ไม่ใช่คำรับประกัน และมันล้มเหลวกับผู้ป่วยที่ถูกปกป้องน้อยที่สุดอยู่แล้ว

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

คำรับประกันความเป็นส่วนตัวเป็นสัญญากับแต่ละคน ไม่ใช่กับค่าเฉลี่ย

เวลาโมเดล AI ทางการแพทย์ถูกเรียกว่า “รักษาความเป็นส่วนตัว” ข้ออ้างนั้นมักตั้งอยู่บนตัวเลขเฉลี่ยหนึ่งค่า: เมื่อรวมผู้ป่วยทั้งหมดที่ข้อมูลถูกใช้ฝึกโมเดล โอกาสที่จะอนุมานได้ว่าบุคคลใดอยู่ในชุดฝึกดูต่ำ ฟังแล้วสบายใจ แต่ประเด็นเงียบ ๆ และน่ากังวลของงานนี้คือ **ค่าเฉลี่ยอาจเป็นตัวเลขผิดตัวสำหรับคำถามเรื่องความเป็นส่วนตัว**

ความเป็นส่วนตัวไม่ใช่คุณสมบัติของ “คนโดยเฉลี่ย” แต่เป็นคำสัญญาที่ให้กับแต่ละคนว่า การที่ข้อมูลของเขาอยู่ในชุดฝึกจะไม่ย้อนกลับมาเปิดเผยตัวเขา ค่าเฉลี่ยอาจดูดีมาก แม้คำสัญญานั้นจะล้มเหลวเกือบสมบูรณ์กับคนเพียงไม่กี่คนก็ตาม

นักวิจัยจึงวัดความเสี่ยงด้านความเป็นส่วนตัว **รายผู้ป่วย** แทนการเฉลี่ยทั้งชุดข้อมูล และพบรูปแบบที่ชัดเจน: โมเดลที่ดูปลอดภัยเมื่อมองภาพรวม อาจเปิดเผยการเป็นสมาชิกของบุคคลบางคนได้เกือบสมบูรณ์ และความเสี่ยงนี้กระจายไม่เท่ากัน กลุ่มที่มีตัวแทนน้อยในข้อมูลมักรับความเสี่ยงมากกว่า

ผู้วิจัยทำอะไร

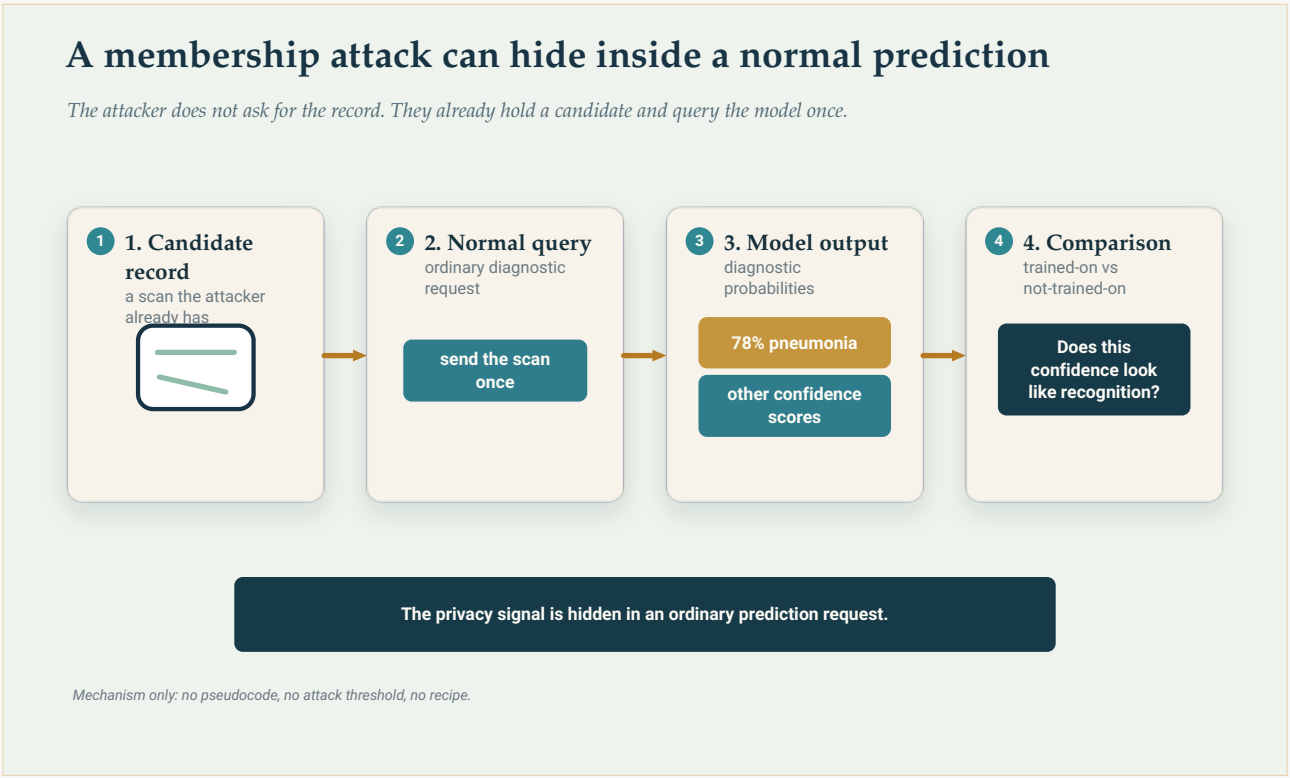
ทีมที่นำโดย Moritz Knolle ร่วมกับ Daniel Rückert จาก Technical University of Munich, Georg Kaissis จาก Hasso Plattner Institute และเพื่อนร่วมงานจาก Imperial College London เริ่มจากภัยคุกคามที่รู้จักกันดีชื่อ **membership inference attack** แล้วเปลี่ยนคำถาม

แทนที่จะถามว่า “โดยเฉลี่ย การโจมตีนี้สำเร็จบ่อยแค่ไหนในทั้งชุดข้อมูล?” พวกเขาถามว่า **“สำหรับผู้ป่วยคนนี้โดยเฉพาะ ความเสี่ยงสูงแค่ไหน?”**

ทีมวิเคราะห์ข้อมูลระดับรายบุคคลใน **ชุดข้อมูลทางการแพทย์มาตรฐาน 7 ชุด** ครอบคลุมภาพถ่ายทางการแพทย์ คลื่นไฟฟ้าหัวใจ และเวชระเบียนอิเล็กทรอนิกส์ โดยทดสอบโมเดล 200 ตัวต่อชุดข้อมูล สำหรับผู้ป่วยแต่ละคนในแต่ละโมเดล พวกเขาประมาณว่าผู้โจมตีสามารถบอกได้มั่นใจเพียงใดว่าข้อมูลของคนนั้นเคยอยู่ในชุดฝึก จากนั้นจึงแยกผลตามกลุ่ม เช่น สถานะโรค เชื้อชาติที่รายงานเอง ประกัน เพศ และวิธีถ่ายภาพ

> **membership inference attack คืออะไร**
>
> การรั่วในที่นี้ละเอียดกว่าภาพที่ว่า “โมเดลคายเวชระเบียนออกมา” สิ่งที่ผู้โจมตีพยายามรู้คือ **membership** — ข้อเท็จจริงเพียงว่าข้อมูลของบุคคลนั้นเคยอยู่ในชุดฝึกหรือไม่
> ลองนึกถึงโมเดลวิเคราะห์ X-ray ปอด เราส่งภาพเข้าไป แล้วโมเดลคืนความน่าจะเป็นของโรคต่าง ๆ (เช่น โอกาสปอดบวม 78%) รวมถึงหัวใจโตหรือภาวะบวมน้ำ การโจมตีใช้เพียงผลลัพธ์ตามปกตินี้ ไม่ต้องมีช่องทางพิเศษ
> จุดอ่อนที่อาศัยคือ โมเดลมักมั่นใจกับตัวอย่างที่เคยเห็นระหว่างฝึกมากกว่าตัวอย่างใหม่เล็กน้อย คล้ายกับนักเรียนที่เคยเห็นข้อสอบมาก่อนและตอบบางข้อได้มั่นใจผิดปกติ การโจมตีพยายามอ่านสัญญาณนี้เพื่อบอกว่า record หนึ่งเคยเป็นตัวอย่างฝึกหรือไม่
> ผู้โจมตีต้องมีตัวอย่างของบุคคลที่ต้องการทดสอบอยู่แล้ว เช่น ภาพสแกนของคนนั้น จากนั้นส่งเข้าโมเดลเหมือนคำขอวินิจฉัยทั่วไป แล้วเปรียบเทียบรูปแบบความมั่นใจกับโมเดลอ้างอิงที่ฝึกเอง เพื่อดูว่าผลลัพธ์คล้ายกรณี “เคยอยู่ในชุดฝึก” หรือ “ไม่เคยอยู่” มากกว่า
> เหตุใดการรู้เพียง membership จึงสำคัญ? เพราะการเป็นสมาชิกของชุดข้อมูลบางชุดอาจเปิดเผยข้อเท็จจริงอ่อนไหวได้ เช่น หากโมเดลถูกฝึกเฉพาะจากผู้ป่วยที่ได้รับการรักษามะเร็งชนิดหนึ่ง การยืนยันว่าข้อมูลของคุณอยู่ในชุดนั้นอาจบอก

> เป็นนัยถึงโรคของคุณ แม้เนื้อหาเวชระเบียนจะไม่เคยถูกเปิดออกมาเลย
> ผู้เขียนระบุสมมติฐานของการโจมตีอย่างตรงไปตรงมา: ผู้โจมตีเข้าถึงผลทำนายตามปกติของโมเดล มี record ผู้สมัครของบุคคลเป้าหมาย และสามารถสร้างโมเดลอ้างอิงของตัวเองได้ งานไม่ได้อ้างว่านี่เป็นการโจมตีที่เกิดขึ้นกับระบบจริงทุกวัน แต่ใช้สถานการณ์นี้เพื่อวัดว่าความเสี่ยงอาจสูงเพียงใด



การโจมตีไม่ต้องใช้ช่องทางความเป็นส่วนตัวพิเศษ คำขอวินิจฉัยตามปกติอาจรั่วสัญญาณว่าข้อมูลของบุคคลนั้นเคยอยู่ในชุดฝึก หากโมเดลตอบด้วยรูปแบบความมั่นใจที่ต่างจากข้อมูลที่ไม่เคยเห็น

Original Aurora diagram — The Clean Paper CC BY 4.0

พบอะไร

มีสามผลหลัก และแต่ละผลทำให้ภาพชัดขึ้น

ค่าเฉลี่ยซ่อนผู้ป่วยที่เสี่ยงสูง ดังที่ Knolle อธิบายไว้ใน ประกาศของงานวิจัย เมื่อวัดแบบรวมทั้งชุดข้อมูล โมเดลจำนวนมากดูปลอดภัย: การโจมตีแทบไม่ดีกว่าการเดาสุ่มสำหรับผู้ป่วยส่วนใหญ่ แต่เมื่อดูรายคน พบว่าบางคนสามารถถูกระบุได้ **เกือบสมบูรณ์** โดยมีค่า AUC ของการโจมตี **0.95 หรือสูงกว่า** ทั้งที่ค่าเฉลี่ยของชุดข้อมูลดูใกล้ระดับสุ่ม (AUC 0.5 หมายถึงแยกไม่ได้ดีกว่าการโยนเหรียญ ส่วน 1.0 คือแยกได้สมบูรณ์)



ค่าเฉลี่ยทั้งชุดข้อมูลอาจอยู่ใกล้ระดับเดาสุ่ม ขณะที่ผู้ป่วยกลุ่มเล็กบริเวณปลายหางของการกระจายยังถูกระบุได้ง่ายมาก ประเด็นของงานคือความเป็นส่วนตัวต้องตรวจในระดับรายบุคคล ไม่ใช่ดูเพียงค่าเฉลี่ย

Original Aurora diagram — The Clean Paper CC BY 4.0

ความเสี่ยงไม่ได้กระจายเท่ากัน และมักตกกับกลุ่มที่มีตัวแทนน้อย ผู้ป่วยที่เสี่ยงต่อการถูกระบุสูงสุดมักมาจากกลุ่มที่มีข้อมูลน้อย เช่น ชนกลุ่มน้อยทางชาติพันธุ์ ผู้ป่วยโรคหายาก หรือภาพถ่ายที่มีลักษณะผิดปกติ ในชุดเวชระเบียนหนึ่ง ผู้ป่วย Black ปรากฏในกลุ่มที่เสี่ยงสูงสุด **มากกว่าที่คาด 31%** ส่วนในชุด mammography ภาพที่ถูกจัดว่าน่าสงสัยต่อมะเร็งมีการกระจุกตัวในกลุ่มเสี่ยงสูงมากถึง **+1,179%**

ตัวเลขเหล่านี้เป็นการ **มีตัวแทนเกินสัดส่วนภายในกลุ่มเสี่ยงสูงสุด** ไม่ได้หมายความว่าผู้ป่วย Black หรือภาพ mammography ที่น่าสงสัยทั้งหมดถูกเปิดเผย ประเด็นที่ทำซ้ำได้คือ กลุ่มที่พบได้น้อยในชุดฝึกมักมีตัวอย่างคล้ายกันน้อย จึงโดดเด่นกว่า และความโดดเด่นนี้เองที่การโจมตีแบบ membership inference ใช้ประโยชน์

โมเดลใหญ่ขึ้นทำให้ปัญหาแย่ลง จำนวนผู้ป่วยที่ถูกโจมตีได้เกือบสมบูรณ์เพิ่มตามความจุของโมเดล ในชุดข้อมูลผิวหนังหนึ่งชุด สัดส่วนนี้เพิ่มจากแทบศูนย์ในโมเดลเล็กที่สุดเป็นประมาณ **หนึ่งในสิบ** ในโมเดลใหญ่ที่สุด ดังนั้นทิศทางของวงการที่สร้างโมเดลใหญ่และเก่งขึ้นไม่ได้รับประกันว่าความเสี่ยงชนิดนี้จะลดลง

Daniel Rückert สรุปประเด็นตรง ๆ ว่า ความเสี่ยงแบบนี้ยอมรับไม่ได้ง่าย ๆ เพราะข้อมูลสุขภาพมีความอ่อนไหวสูง

ทำไม “ปลอดภัยโดยเฉลี่ย” จึงเป็นการทดสอบผิดแบบ

สิ่งล่อใจคืออ่านความเสี่ยงเฉลี่ยต้ำว่าเป็นใบรับรองว่าความเป็นส่วนตัวดีพอ แต่บทเรียนแกนกลางของงานคือ **การเฉลี่ยวัดผิดคำถาม**

คำรับประกันด้านความเป็นส่วนตัวมีความหมายก็ต่อเมื่อถือกับคนที่เสี่ยงที่สุด ไม่ใช่เพียงคนส่วนใหญ่ โมเดลที่ผู้ป่วย 999 จาก

1,000 คนถูกระบุไม่ได้ แต่คนหนึ่งถูกระบุได้เกือบสมบูรณ์ ไม่ได้ “ปลอดภัย 99.9%” ในความหมายที่สำคัญต่อคนคนนั้น และหากคนที่เสี่ยงสูงมีแนวโน้มเป็นผู้ป่วยโรคหายากหรือสมาชิกของชนกลุ่มน้อย ตัวชี้วัดเฉลี่ยไม่ได้เพียงหยาบ แต่ยังกลบความไม่เท่าเทียมด้วย

งานนี้จึงบังคับให้เปลี่ยนคำถามจาก

โมเดลนี้รักษาความเป็นส่วนตัวได้ดีเพียงใดโดยเฉลี่ย?

เป็น

ผู้ป่วยที่เสี่ยงถูกเปิดเผยที่สุดคือใคร และคนเหล่านั้นมาจากกลุ่มใด?

คำถามที่สองต่างหากที่ตรงกับความหมายของคำว่า “ความเป็นส่วนตัว” มากกว่า

สิ่งที่ผลนี้ *ไม่ได้* พิสูจน์

- **ไม่ได้บอกให้เลิกใช้ AI ทางการแพทย์** กรอบของผู้เขียนคือการวัดและลดความเสี่ยง ไม่ใช่หยุดพัฒนา
- **ไม่ได้แสดงว่าโมเดลทางการแพทย์ทุกตัวกำลังรั่วข้อมูล** หรือว่าระบบที่ใช้งานจริงตัวใดถูกโจมตีแล้ว งานแสดงว่าความเสี่ยงนี้มีอยู่และกระจายไม่เท่ากัน ภายใต้สมมติฐานการโจมตีที่กำหนด
- **ไม่ได้หมายความว่าเวชระเบียนของคุณเปิดอยู่แล้ว** การโจมตีต้องเข้าถึงโมเดล มี record ผู้สมัคร และมีโครงสร้างสำหรับเปรียบเทียบ ไม่ใช่ความสามารถที่ใครก็ทำได้โดยบังเอิญ
- **ไม่ได้แสดงว่าเนื้อหาเวชระเบียนรั่วออกมา** สิ่งที่ถูกอนุมานคือการเป็นสมาชิกของชุดฝึก ซึ่งอาจอ่อนไหวด้วยเหตุผลของมันเอง
- **ไม่ได้สรุปความเหลื่อมล้ำทั้งหมดด้วยตัวเลขเดียว** ผลที่แข็งแรงคือรูปแบบ: ค่าเฉลี่ยประเมินความเสี่ยงรายบุคคลต่ำเกินไป และกลุ่มที่มีตัวแทนน้อยรับความเสี่ยงมากกว่าในหลายชุดข้อมูล

หลักฐานน่าเชื่อถือเพียงใด

นี่เป็นผลเชิงประจักษ์และเชิงวิธีวิทยาที่ค่อนข้างแข็งแรง รูปแบบหลัก — ตัวชี้วัดแบบรวมประเมินความเสี่ยงรายบุคคลต่ำไป ความเสี่ยงที่เหลือกระจุกในกลุ่มที่มีตัวแทนน้อย และปัญหาแย่ลงเมื่อโมเดลใหญ่ขึ้น — ปรากฏใน **ชุดข้อมูลทางการแพทย์ 7 ชุดที่มีชนิดข้อมูลต่างกัน** และโมเดลจำนวนมากต่อชุด ความกว้างนี้ทำให้ข้อสรุปเรื่องวิธีวัดน่าเชื่อมากกว่าการพบในชุดข้อมูลเดียว

มีข้อควรระวังสองข้อ

อย่างแรก นี่เป็นการสาธิต **ความเสี่ยงที่เป็นไปได้** จากการโจมตีที่ผู้เขียนสร้างและทดสอบเอง บอกว่าผู้โจมตีที่มีความสามารถอาจทำได้ดีเพียงใดภายใต้สมมติฐานที่ระบุ ไม่ได้บอกว่าการโจมตีชนิดนี้เกิดในโลกจริงบ่อยเพียงใด

อย่างที่สอง ตัวเลขที่แรงที่สุด เช่น AUC ใกล้ 1 ถูกเลือกมาเพื่ออธิบาย **คนที่เสี่ยงสูงที่สุด** โดยตั้งใจ นั่นคือประเด็นของงาน และควรอ่านว่า “ปลายหางของความเสี่ยงหนักกว่าที่ค่าเฉลี่ยบอกมาก” ไม่ใช่ “ผู้ป่วยส่วนใหญ่ถูกเปิดเผย”

ท่าทีที่เหมาะจึงไม่ใช่ตื่นตระหนกหรือปัดทิ้ง แต่คือ: งานหลายชุดข้อมูลที่ระมัดระวังแสดงว่า วิธีมาตรฐานที่ใช้บอกว่า AI ทางการแพทย์รักษาความเป็นส่วนตัวได้ อาจมองไม่เห็นกรณีเลวร้ายที่สุด — และกรณีเหล่านั้นมักตกกับผู้ป่วยที่มีพื้นที่ป้องกันน้อยที่สุดอยู่แล้ว

ทำไมจึงสำคัญ

มีสองเหตุผลที่ทำให้งานนี้มากกว่ารายละเอียดทางเทคนิค

อย่างแรก มันเปลี่ยนว่าคำว่า **"privacy-preserving"** ควรได้รับอนุญาตให้หมายถึงอะไร หากโมเดลมีคะแนนความเป็นส่วนตัวเฉลี่ยต่ำ คะแนนนั้นอาจจริง แต่ยังซ่อนผู้ป่วยกลุ่มเล็กที่ membership inference ระบุได้ง่ายมาก ผู้เขียนเสนอในทางปฏิบัติว่าควรประเมินความเสี่ยง **รายผู้ป่วย** ก่อนปล่อยโมเดล ควบคุมว่าใครเข้าถึงระบบที่ใช้งานจริงได้ และใช้เทคนิคอย่าง **differential privacy** ซึ่งเติมความสุ่มอย่างปรับระดับระหว่างการฝึกเพื่อทำให้การอนุมาน membership ยากขึ้น โดยยอมรับว่ามีต้นทุนต่อความแม่นยำหรือประโยชน์ของโมเดล

กล่าวอีกอย่าง "เราตรวจค่าเฉลี่ยแล้ว" ไม่ควรถือว่า "ตรวจความเป็นส่วนตัวแล้ว" อีกต่อไป

อย่างที่สอง งานเชื่อม **ความเป็นส่วนตัวเข้ากับความเป็นธรรม** กลุ่มเดียวกันที่มีตัวแทนน้อยในข้อมูลทางการแพทย์ — และจึงอาจได้รับบริการจาก AI แย่กว่าอยู่แล้ว — กลับเป็นกลุ่มที่ได้รับการปกป้องด้านความเป็นส่วนตัวน้อยลงด้วย วงการที่กำลังพยายามแก้ความเหลื่อมล้ำด้านประสิทธิภาพจึงไม่สามารถมองความเหลื่อมล้ำด้าน privacy เป็นปัญหาคนละแผนกได้

เรื่องปลอบใจเดิมว่า *"เราวัดแล้ว ค่าเฉลี่ยต่ำ ทุกอย่างโอเค"* คือสิ่งที่งานนี้รื้อออก ไม่ใช่เพื่อทำให้คนกลัวเทคโนโลยี แต่เพื่อย้ายมาตรฐานไปยังจุดที่ควรอยู่: **คำสัญญาว่าปกป้องทุกคนต้องตรวจให้ถึงคนที่มีโอกาสเสียหายมากที่สุดด้วย**

สรุป

Membership inference attack พยายามบอกว่า record ของบุคคลหนึ่งเคยอยู่ในข้อมูลฝึกของโมเดลหรือไม่ และเพราะข้อเท็จจริงเรื่องการเป็นสมาชิกอาจเปิดเผยข้อมูลสุขภาพ การรั่วแบบนี้จึงมีความหมายแม้เนื้อหาเวชระเบียนจะไม่เคยถูกแสดงออกมา งานก่อนหน้ามักวัดความสำเร็จของการโจมตีเฉลี่ยทั้งชุดข้อมูล งานนี้วัด **รายผู้ป่วย** ในชุดข้อมูลทางการแพทย์ 7 ชุด ครอบคลุมภาพถ่ายทางการแพทย์ ECG และเวชระเบียนอิเล็กทรอนิกส์ โดยใช้โมเดลจำนวนมากต่อชุด พบว่าค่าเฉลี่ยประเมินความเสี่ยงต่ำไปมากสำหรับบางคน บุคคลบางรายถูกระบุได้เกือบสมบูรณ์ (AUC $\geq$ 0.95) แม้ค่าเฉลี่ยของชุดข้อมูลดูปลอดภัย กลุ่มที่เสี่ยงสูงสุดมักเป็นกลุ่มที่มีตัวแทนน้อย เช่น ชนกลุ่มน้อย โรคหายาก หรือภาพผิดปกติ และปัญหาเพิ่มตามความจุของโมเดล ผู้เขียนไม่ได้เสนอให้เลิกใช้ AI ทางการแพทย์ แต่เสนอให้วัดความเสี่ยงระดับบุคคล จำกัดการเข้าถึงโมเดล และใช้เทคนิคอย่าง differential privacy ข้อสรุปคือ: **"ปลอดภัยโดยเฉลี่ย" ไม่ใช่คำรับประกันความเป็นส่วนตัว และคนที่ค่าเฉลี่ยมองข้ามมักเป็นคนที่ได้รับการปกป้องน้อยที่สุดอยู่แล้ว**

ประเมินแบบไม่เกินจริง

งานวิจัยแสดงอะไร: ในชุดข้อมูลทางการแพทย์ 7 ชุดและโมเดลจำนวนมาก ความเสี่ยง membership inference ของผู้ป่วยบางรายสูงกว่าที่ตัวชี้วัดเฉลี่ยบอกมาก ถึงระดับที่การโจมตีเกือบแยกได้สมบูรณ์ (AUC $\geq$ 0.95) ความเสี่ยงที่เหลือกระจุกในกลุ่มที่มีตัวแทนน้อยและเพิ่มตามความจุของโมเดล

อะไรที่เป็นไปได้แต่ยังไม่พิสูจน์: ผู้โจมตีในโลกจริงอาจใช้ความสามารถนี้กับระบบที่ใช้งานอยู่ งานนี้วัดความเป็นไปได้และการกระจายของความเสี่ยงภายใต้สมมติฐานที่ระบุ ไม่ได้วัดความถี่ของเหตุการณ์โจมตีจริง

งานไม่ได้แสดงอะไร: ไม่ได้บอกให้เลิก AI ทางการแพทย์ ไม่ได้แสดงว่าโมเดลทุกตัวรั่วหรือระบบจริงตัวใดถูกเจาะ ไม่ได้แสดงว่าเนื้อหาเวชระเบียนถูกเปิด และไม่ได้ย่อความเหลื่อมล้ำทั้งหมดเป็นตัวเลขเดียว

ข้อจำกัดหลัก: เป็นการวัดความเสี่ยงด้วยการโจมตีที่ผู้เขียนสร้างขึ้น ไม่ใช่เหตุการณ์ที่สังเกตในโลกจริง ตัวเลขที่แรงที่สุดจงใจอธิบายผู้ป่วยปลายหางที่เสี่ยงสูง และรูปแบบความเหลื่อมล้ำแม้เกิดซ้ำหลายชุดข้อมูลก็ไม่ควรถูกอ่านเป็นค่าคงที่เดียวสำหรับทุกบริบท

ผู้อ่านทั่วไปควรมั่นใจแค่ไหน? มั่นใจสูงว่าตัวชี้วัดความเป็นส่วนตัวแบบเฉลี่ยสามารถประเมินความเสี่ยงรายบุคคลต่ำเกินไป ความเสี่ยงที่เหลือมักกระจุกในผู้ป่วยที่มีตัวแทนน้อย และปัญหามีแนวโน้มเพิ่มเมื่อโมเดลใหญ่ขึ้น ส่วนความถี่ของการโจมตีจริงควรมีความมั่นใจต่ำกว่าเพราะงานนี้ไม่ได้วัดโดยตรง การอ่านที่เหมาะคือ: **วิธีตรวจ privacy แบบเดิมมองไม่เห็นกรณีเลวร้ายที่สุดของตัวเอง และกรณีเหล่านั้นมักตกกับผู้ป่วยที่เปราะบางที่สุด — ดังนั้นวิธีประเมินต้องเปลี่ยน**

แหล่งข้อมูล

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