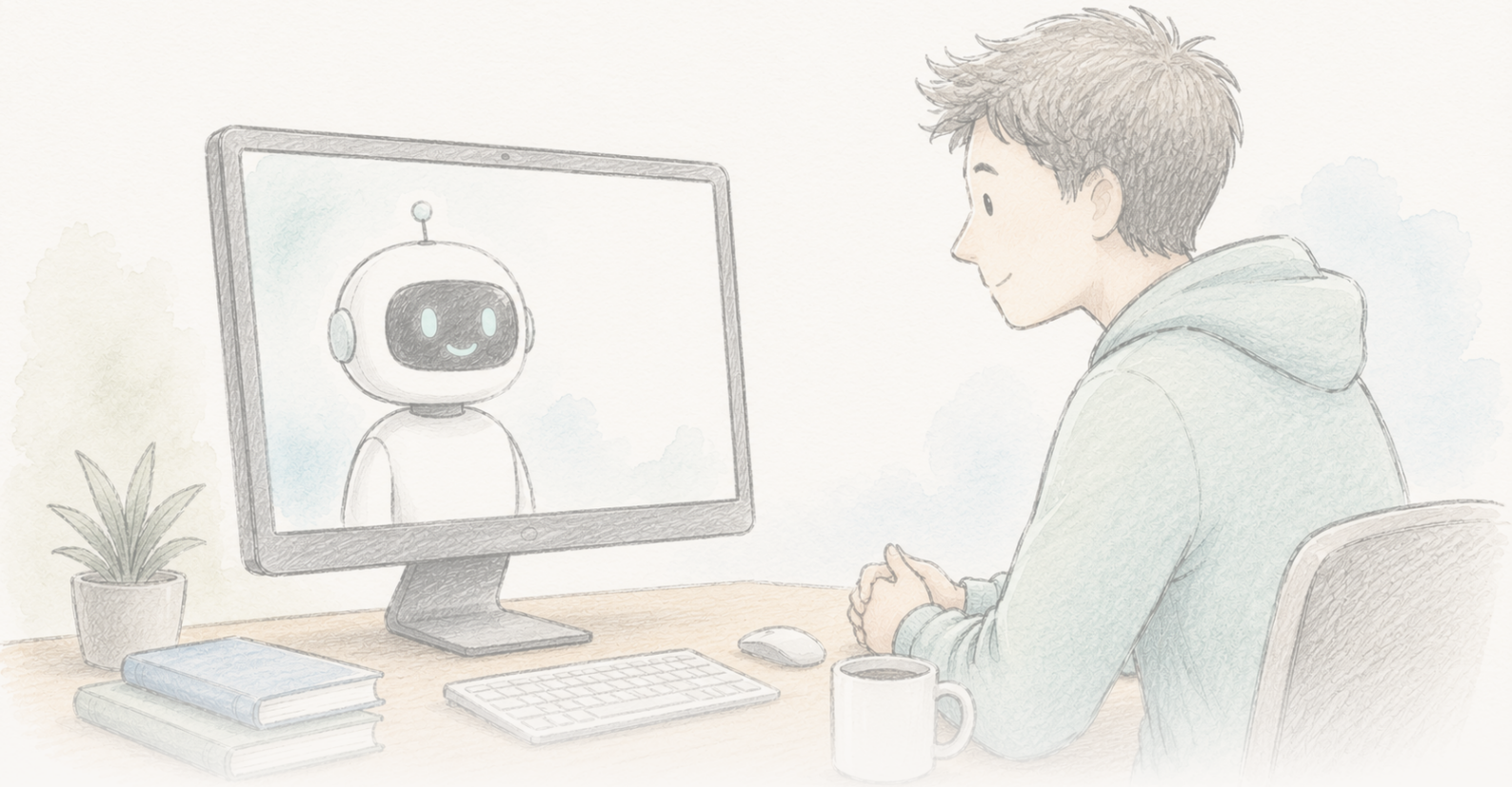


The Clean Paper

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

Chatbot ดูเหมือนจะลดความเชื่อทฤษฎีสมคบคิดได้นานหลายเดือน

งานวิจัยปี 2024 พบว่าการสนทนาสามรอบกับ GPT-4 Turbo ลดความเชื่อของผู้คนในทฤษฎีสมคบคิดที่ตนเชื่ออยู่ราว 20% ผลคงอยู่นานสองเดือน พบในหลายประเภทของทฤษฎี และตั้งอยู่บนข้อกล่าวอ้างจาก AI ที่ถูกประเมินว่าถูกต้องเกือบทั้งหมด เดือนมิถุนายน 2026 บทความถูกติด Editorial Expression of Concern จากปัญหาการจัดการข้อมูลและ reproducibility; ผู้เขียนบอกว่าการวิเคราะห์ที่แก้แล้วยังคงผลทั้งทิศทาง นัยสำคัญ และขนาด และ Science ยังประเมินอยู่ เป็นผลที่น่าสนใจจริงและออกแบบอย่างระมัดระวัง — แต่ขณะนี้อยู่ระหว่างตรวจสอบ ยังไม่ยุติ และยังไม่ถูกหักล้าง

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

Paper: *Durably reducing conspiracy beliefs through dialogues with AI*, Thomas H. Costello, Gordon Pennycook, David G. Rand, Science 385, eadq1814 (2024) · DOI: 10.1126/science.adq1814

Editorial Expression of Concern

Science ติด Editorial Expression of Concern ให้บทความนี้ระหว่างประเมินคำถามเรื่องการจัดการข้อมูลและ reproducibility นี่ไม่ใช่ retraction และไม่ใช่ข้อสรุปว่ามี misconduct; หมายความว่าควรอ่านผลที่รายงานเป็น provisional จนกว่าการตรวจสอบจะจบ

Editorial Expression of Concern →

สุดท้ายก็ยังเป็นเรื่องของข้อเท็จจริง — และตอนนี้งานวิจัยมีธงเตือนติดอยู่

ในปี 2024 มีงานวิจัยรายงานผลที่สวนทางกับความเชื่อสบาย ๆ อย่างหนึ่ง: ถ้าให้คนคุยสั้น ๆ สามรอบกับ AI — GPT-4 Turbo ที่ถูก พรอมป์ต์ ให้ตอบหลักฐานเฉพาะที่คนนั้นยกขึ้นมาสนับสนุนทฤษฎีสมคบคิดที่เขาเชื่อจริง — โดยเฉลี่ยความเชื่อในทฤษฎีนั้นลดลงประมาณ 20% และการลดลงยังเห็นได้อีกสองเดือนต่อมา ผลนี้เกิดทั้งกับทฤษฎีสมคบคิดคลาสสิก (การลอบสังหาร John F. Kennedy มนุษย์ต่างดาว Illuminati) และเรื่องร่วมสมัย (COVID-19 การเลือกตั้งสหรัฐฯ ปี 2020) รวมถึงในคนที่เชื่อแรงและผูกความเชื่อนั้นกับอัตลักษณ์ของตัวเอง เมื่อนักตรวจสอบข้อเท็จจริงมืออาชีพให้คะแนนข้อกล่าวอ้าง 128 ข้อที่ AI พูด 127 ข้อเป็นจริง 1 ข้อชวนเข้าใจผิด และไม่มีข้อใดเป็นเท็จ

พาดหัวที่แพร่ไปคือ คุณ *สามารถ* คุยคนให้ออกจาก rabbit hole ได้ — ผู้เชื่อทฤษฎีสมคบคิดไม่ได้อยู่นอกขอบเขตของหลักฐาน เพียงแต่ต้องได้รับหลักฐานที่ตรงกับเหตุผลเฉพาะของตัวเองมากพอ นี่เป็นข้ออ้างที่น่าสนใจจริง และการศึกษาก็ออกแบบอย่างระมัดระวังกว่างานด้าน persuasion จำนวนมาก

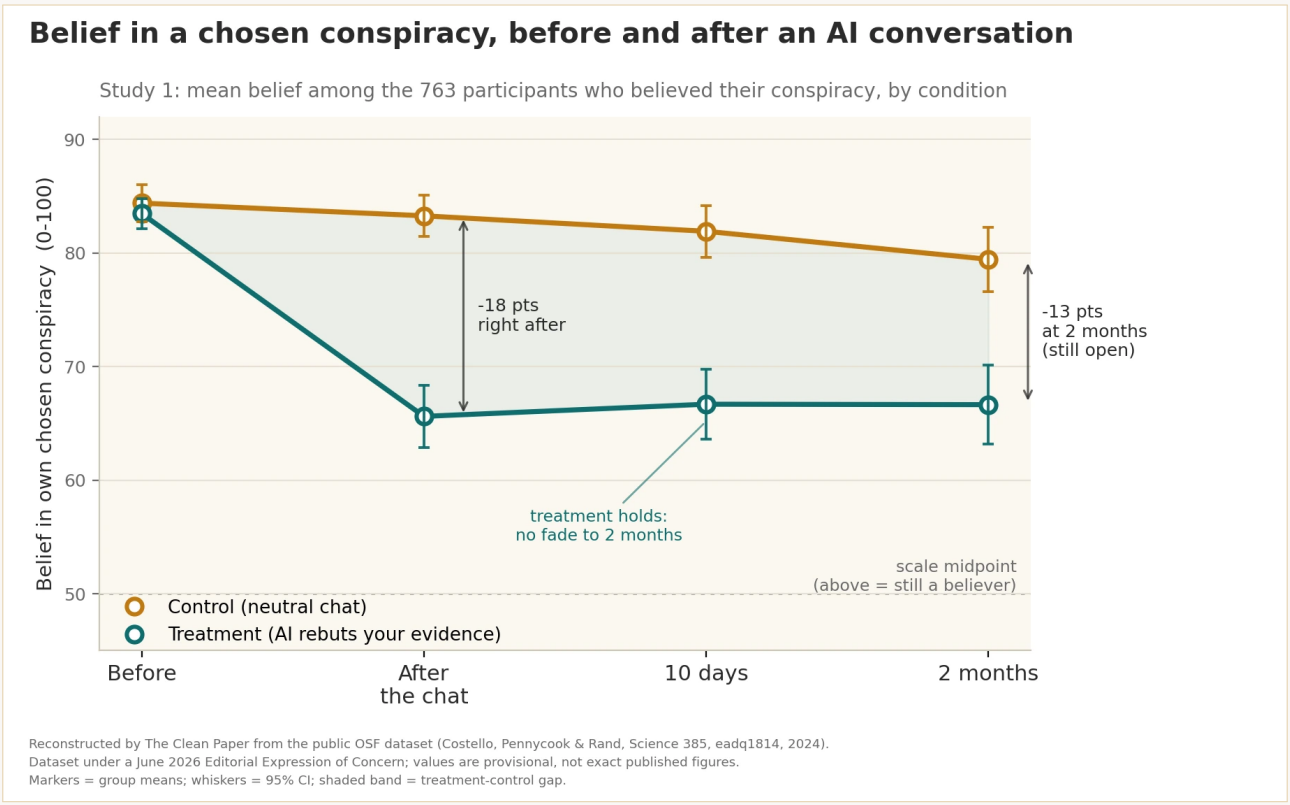
จากนั้นเดือนมิถุนายน 2026 *Science* ติด **Editorial Expression of Concern** ให้กับบทความ ส่วนนี้คือสิ่งที่การเล่าเรื่องซ้ำจำนวนมากน่าจะข้าม และกลับเป็นส่วนที่กำหนดว่าตอนนี้เราควรให้น้ำหนักผลลัพธ์ได้มากแค่ไหน

Expression of Concern คืออะไร — และไม่ใช่อะไร

Editorial Expression of Concern เป็นธงเตือนอย่างเป็นทางการที่วารสารแนบกับบทความเพื่อบอกผู้อ่านว่ามีคำถามสำคัญถูกยกขึ้นมา ขณะที่การตรวจสอบยังไม่เสร็จ มัน **ไม่ใช่** retraction (บทความไม่ได้ถูกถอน) และตัวมันเองก็ไม่ใช่ข้อสรุปว่ามี misconduct ความหมายคือ: อ่านงานนี้โดยมีข้อสงวนไว้ เพราะบันทึกทางวิทยาศาสตร์อาจเปลี่ยนได้ ในกรณีนี้ หลังได้รับแจ้งปัญหา ผู้เขียนตรวจสอบเอง รายงานรายละเอียดต่อ *Science* และส่งการวิเคราะห์ที่แก้ไขแล้ว

สิ่งที่ถูกตั้งธงมีรายละเอียดชัด ผู้เขียนได้รับแจ้งว่ามีความไม่สอดคล้องกันระหว่างเกณฑ์คัดกรองผู้เข้าร่วมที่อธิบายใน manuscript กับเกณฑ์ที่ใช้จริงในโค้ดวิเคราะห์ที่เผยแพร่ และชุดข้อมูลสาธารณะมีแถวส่วนเกินที่ถูกสอดเข้ามาจากข้อผิดพลาดตอนรวมโค้ด — ปัญหาที่ทำให้ตัวเลขบางส่วนในบทความทำซ้ำจากวัสดุที่ปล่อยออกมาได้ยาก ผู้เขียนส่ง pipeline การวิเคราะห์ที่แก้ไขแล้วและผลอัปเดตให้ *Science* ซึ่งพวกเขาบอกว่ายังคงเหมือนเดิมทั้ง **ทิศทาง นัยสำคัญทางสถิติ และขนาดเชิงสาระ** *Science* กำลังประเมินอยู่ จนกว่าการประเมินนั้นจะเสร็จ ตัวเลขที่แน่นอนยังมีเครื่องหมายคำถามซึ่งมีเพียงวารสารเท่านั้นที่ยกออกได้

ดังนั้นนี่คือสองเรื่องพร้อมกัน: ผลที่น่าสนใจเกี่ยวกับว่าข้อเท็จจริงสามารถขยับความเชื่อที่ฝังแน่นได้หรือไม่ และตัวอย่างสด ๆ ของบันทึกวิทยาศาสตร์ที่กำลังแก้ไขตัวเองอย่างเปิดเผย เป้าหมายของบทความนี้คือไม่เอาสองเรื่องนั้นมาปนกัน



ในการศึกษา การคุยกับ AI สามรอบที่ตอบโต้หลักฐานซึ่งผู้เข้าร่วมยกขึ้นมาเอง ถูกรายงานว่าลดความเชื่อในทฤษฎีสมคบคิดที่คนนั้นเลือกลงประมาณ 20% โดยเฉลี่ย; ต่างจากกลุ่มควบคุมที่คุยเรื่องไม่เกี่ยวข้อง การลดลงยังวัดได้ที่ 10 วันและ 2 เดือน ผลที่รายงานจึงคงอยู่แต่ไม่สมบูรณ์: ผู้เข้าร่วมส่วนใหญ่ยังเชื่อทฤษฎีนั้นหลังการสนทนา — เป็นการลด ไม่ใช่การรักษาให้หาย งานนี้อยู่ภายใต้ Editorial Expression of Concern (*Science*, มิถุนายน 2026) จากความไม่สอดคล้องในการรายงานและข้อผิดพลาดในชุดข้อมูลสาธารณะ; ผู้เขียนบอกว่าการวิเคราะห์ที่แก้แล้วยังคงทิศทาง นัยสำคัญ และขนาดผล ขณะที่ *Science* ยังประเมินอยู่ กราฟนี้ The Clean Paper สร้างใหม่จากชุดข้อมูลสาธารณะนั้น ดังนั้นค่าที่แสดงยังเป็นค่าชั่วคราว ไม่ใช่ตัวเลขสุดท้ายของบทความ

Original figure — The Clean Paper CC BY 4.0

ผู้วิจัยทำอะไร

- ทำการทดลองสองชุดกับผู้เข้าร่วม 2,190 คน แต่ละคนระบุด้วยคำของตัวเองว่าตนเชื่อทฤษฎีสมคบคิดเรื่องใด และคิดว่ามีหลักฐานอะไรสนับสนุน
- ให้ผู้เข้าร่วมแต่ละคนสนทนากับ GPT-4 Turbo สามรอบ ในเงื่อนไข **treatment** AI ถูก พรอมป์ต์ ให้โต้หลักฐานเฉพาะของผู้เข้าร่วม ส่วนเงื่อนไข **control** คุยเรื่องที่ไม่เกี่ยวข้อง
- วัดความเชื่อในทฤษฎีสมคบคิดที่เลือกก่อนและทันทีหลังการสนทนา แล้วติดต่อกลับที่ **10 วันและ 2 เดือน**
- ให้นักตรวจสอบข้อเท็จจริงมืออาชีพประเมินความถูกต้องของข้อกล่าวอ้างเชิงข้อเท็จจริงทั้ง 128 ข้อที่ AI พูดในตัวอย่างหนึ่ง
- ตรวจว่าผลลัพธ์ลามไปยังทฤษฎีสมคบคิดอื่นที่ไม่เกี่ยวข้องหรือไม่ — และเพื่อทดสอบความจำเพาะ ยังตรวจด้วยว่ามันลดความเชื่อใน “สมคบคิด” ที่บังเอิญ **เป็นเรื่องจริง** หรือไม่

พวกเขาพบอะไร

- **ความเชื่อลดลงโดยเฉลี่ยประมาณ 20%** ในทฤษฎีสมคบคิดที่เลือกของกลุ่ม treatment เทียบกับ control (การศึกษา 1: ช่วงความเชื่อมั่น 95% [13.8, 19.7] จุดบนสเกล 100 จุด, $P < 0.001$)
- **ผลคงอยู่** ในกลุ่ม treatment การลดลงไม่ได้เด้งกลับ: ความเชื่อยังอยู่ใกล้ระดับหลังคุยทันทีที่ 10 วันและสองเดือน แทนที่จะค่อย ๆ สูงกลับ ขณะที่ control สูงกว่าตลอด บทความอธิบายว่าผลต่อทฤษฎีที่เลือกคงอยู่โดยแทบไม่ลดทอนอย่างน้อยสองเดือน ไม่ใช่แค่การสั่นไหวชั่วคราว
- **ผลเกิดกับหลายประเภท** ของทฤษฎีสมคบคิดที่คนระบุ และยังพบในผู้เข้าร่วมที่เชื่อแรงตั้งแต่ต้น
- **ข้อกล่าวอ้างของ AI มีความถูกต้องสูง** ในบริบทนี้: 127 จาก 128 ข้อ (99.2%) ถูกจัดว่าจริง, 1 ข้อ (0.8%) ชวนเข้าใจผิด, ไม่มีข้อใดเป็นเท็จ
- **ผลมีความจำเพาะ** ไม่ใช่ทำให้สงสัยทุกอย่าง: intervention **ไม่ได้** ลดความเชื่อในสมคบคิดที่เป็นเรื่องจริง จึงชี้ว่ามันขยับความเชื่อที่ไม่มีหลักฐานรองรับ แทนที่จะทำให้คน cynic ต่อทุกเรื่อง
- **มีผลลามเล็กน้อย** — ความเชื่อในทฤษฎีสมคบคิดอื่นที่ไม่เกี่ยวข้องลดลงราว 12% และผู้เข้าร่วมรายงานความตั้งใจสูงขึ้นที่จะโต้ข้อกล่าวอ้างสมคบคิด ผลหลักทำซ้ำใน การศึกษา 2 และชี้ทิศเดียวกันในตัวอย่างเล็กกว่าที่ไม่มีกลุ่มควบคุม (หลักฐานอ่อนกว่า แต่สอดคล้อง)

สิ่งที่ผลนี้ไม่ได้พิสูจน์

- ตอนนี้มัน **ไม่ได้อยู่ในสถานะไร้ข้อกังขา** บทความอยู่ภายใต้ Editorial Expression of Concern จากปัญหาการจัดการข้อมูลและ reproducibility และตัวเลขที่แก้แล้วยังถูก *Science* ประเมินอยู่ ควรมองค่าจำเพาะเป็น provisional จนกว่าการตรวจสอบจะจบ
- มัน **ไม่ได้** แสดงว่า AI “รักษา” ความเชื่อสมคบคิดให้หาย การลดเฉลี่ย ~20% เป็นการเปลี่ยนที่มีความหมาย ไม่ใช่การลบความเชื่อ — ผู้เข้าร่วมส่วนใหญ่ยังเชื่อทฤษฎีนั้นหลังจากนั้น เพียงแต่เชื่อน้อยลง
- นี่ **ไม่ใช่** ผลจาก deployment ในโลกจริง ผู้เข้าร่วมสมัครเข้าการศึกษาและคุยกับ AI อย่างร่วมมือ; ในโลกจริง คนที่ผูกพันกับทฤษฎีสมคบคิดมากที่สุดอาจเป็นคนที่มีโอกาสน้อยที่สุดที่จะไปหา chatbot เพื่อให้มันโต้ตัวเอง
- ความถูกต้อง 99.2% **จำเพาะต่อภารกิจ โมเดล และ พรอมป์ต์ ที่ออกแบบอย่างระมัดระวังนี้** ไม่ใช่การรับประกันทั่วไปว่า language model จะพูดเรื่องจริง ผู้เขียนระบุชัดว่าพลัง persuasion แบบเฉพาะบุคคลเดียวกันนี้สามารถผลัก *ความเชื่อเท็จ* ได้ หากถูกใช้ไปในทิศนั้น
- “คงทน” ในที่นี้หมายถึง **สองเดือน** ซึ่งน่าประทับใจสำหรับการสนทนาครั้งเดียว แต่ไม่เท่ากับถาวร

หลักฐานแข็งแรงแค่ไหน

- **การออกแบบเป็นจุดแข็งจริง** มีการทดลอง treatment เทียบ control, control แบบ ยาหลอก ที่สะอาด, การทำซ้ำสองการศึกษาและตัวอย่างอิสระ, การติดตามความคงทน และการตรวจความถูกต้องของข้อกล่าวอ้างจาก AI โดยตรง — ระมัดระวังกว่างาน persuasion ส่วนใหญ่ และทิศทางผลสอดคล้องทุกที่ที่ทดสอบ
- **แต่ Expression of Concern ไม่ใช่เชิงอรรถเล็ก ๆ** ความสามารถในการสร้างตัวเลขที่รายงานขึ้นใหม่จากข้อมูลและโค้ดที่เผยแพร่เป็นส่วนหนึ่งของความน่าเชื่อถือ และตรงนี้เองที่เกิดปัญหา — เกณฑ์คัดกรองไม่สอดคล้องกันและมีแถวซ้ำ/ส่วนเกินจากข้อผิดพลาดตอน merge โค้ด คำกล่าวของผู้เขียนว่า pipeline ที่แก้แล้วยังคงผลเดิมเป็นสัญญาณที่ให้กำลังใจและเป็นไปได้ แต่ยังเป็นคำอธิบายจากผู้เขียนเอง ไม่ใช่คำตัดสินอิสระ การประเมินของ *Science* คือสิ่งที่ต้องรอ
- **สถานะที่ซื่อสัตย์คือ “ถูกตั้งธง” ไม่ใช่ “หักล้างแล้ว” หรือ “ยืนยันแล้ว”** เป็นงานที่ออกแบบดีและผลสะดุดตา แต่

ตัวเลขแน่นอนอยู่ระหว่างการทบทวนอย่างเป็นทางการ การปล่อยเรื่องดี ๆ ไว้ในสถานะไม่สบายใจแบบนี้อาจไม่น่าพอใจ แต่นี่คือสถานะที่ถูกต้อง

ทำไมจึงสำคัญ

หลายปีมานี้มีมุมมองที่ทรงอิทธิพลว่า ผู้เชื่อทฤษฎีสมคบคิดขับเคลื่อนด้วยความต้องการทางจิตวิทยาและอัตลักษณ์ จึงไม่สามารถใช้ข้อเท็จจริงโต้จนเปลี่ยนใจได้ หากผลลดความเชื่อที่คงอยู่และตั้งอยู่บนหลักฐานนี้ยืนได้จริง มันจะเป็นน้ำหนักถ่วงสำคัญต่อความมองโลกในแง่ร้ายนั้น: ปัญหาอาจอยู่ส่วนหนึ่งที่การ debunk แบบทั่วไปไม่เคยตอบ *หลักฐานเฉพาะ* ที่ผู้เชื่อยกขึ้นมาจริง ซึ่ง LLM สามารถทำในระดับใหญ่ได้

เรื่องนี้ยังสำคัญในฐานะกรณีศึกษาว่าวิทยาศาสตร์ควรทำงานอย่างไร บทเรียนจาก Expression of Concern ไม่ใช่ว่าผลที่โด่งดังไม่มีค่า แต่คือข้อผิดพลาดถูกเปิดออก ข้อมูลถูกตรวจใหม่ และคำกล่าวถูกเช็กซ้ำต่อสาธารณะ กลไกแบบนี้ทำงานอยู่เป็นคุณสมบัติ ไม่ใช่เรื่องอื้อฉาว — และเป็นเหตุผลว่าท่าทีที่เหมาะสมต่อผลนี้คือความอดทน ไม่ใช่ทั้งการปรบมือทันทีหรือปัดทิ้งทันที

และในเรื่อง AI มันคมได้สองด้าน ความสามารถเดียวกันที่ลดความเชื่อเท็จด้วยข้อโต้แย้งที่ปรับให้ตรงกับคน อาจเพิ่มความเชื่อเท็จได้มีประสิทธิภาพพอ ๆ กัน เทคนิคเป็นกลาง แต่เป้าหมายไม่เป็นกลาง

สรุป

งานวิจัยปี 2024 พบว่าการสนทนาสามรอบกับ GPT-4 Turbo ลดความเชื่อของผู้คนในทฤษฎีสมคบคิดที่ตนเชื่ออยู่ราว 20% ผลยังอยู่สองเดือนและพบในหลายประเภทของทฤษฎี โดยข้อกล่าวอ้างเชิงข้อเท็จจริงของ AI ถูกจัดว่าถูกต้องเกือบทั้งหมด เดือนมิถุนายน 2026 บทความถูกติด Editorial Expression of Concern จากปัญหาการจัดการข้อมูลและ reproducibility; ผู้เขียนรายงานว่าการวิเคราะห์ที่แก้แล้วยังคงผลทั้งทิศทาง นัยสำคัญ และขนาด ขณะที่ *Science* ยังประเมินอยู่ ควรอ่านมันเป็นผลที่น่าสนใจจริงและออกแบบอย่างระมัดระวัง แต่กำลังอยู่ระหว่างตรวจสอบ — ยังไม่ปิดเรื่อง และยังไม่ถูกหักล้าง ประโยคที่ซื่อสัตย์ที่สุดคือสิ่งที่กระบวนการเองกำลังบอก: ตรวจอีกครั้ง

แหล่งข้อมูล

Costello, Pennycook & Rand, Durably reducing conspiracy beliefs through dialogues with AI, Science 385, eadq1814 (2024)
`https://doi.org/10.1126/science.adq1814`

Editorial Expression of Concern, Science (posted 11 June 2026; H. Holden Thorp, Editor-in-Chief), DOI 10.1126/science.aej2383
`https://www.science.org/doi/10.1126/science.aej2383`