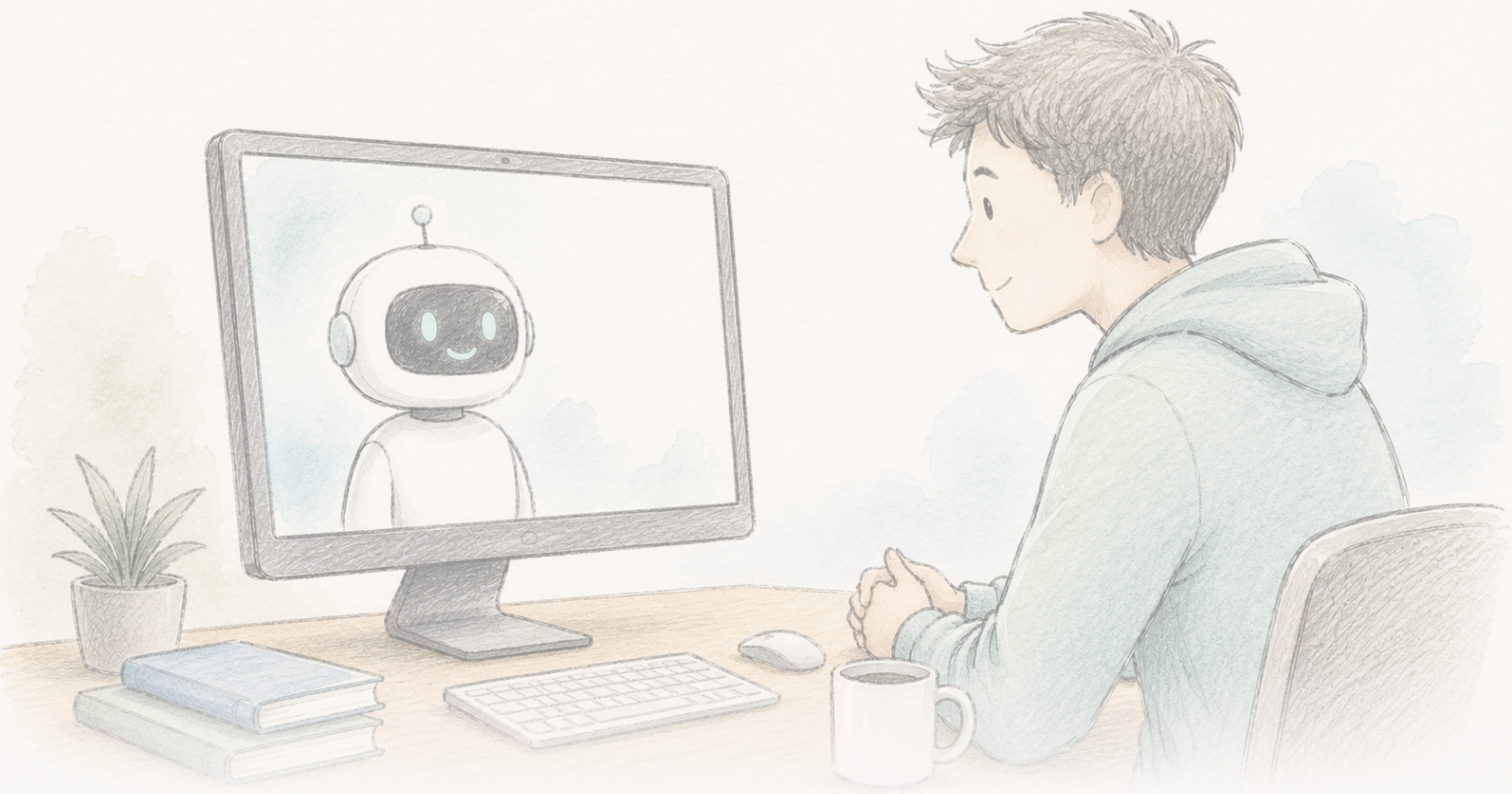




The CLEAN PAPER

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

Chatbot-এর সঙ্গে কথোপকথন conspiracy belief কয়কে মাস কমিয়ে দিচ্ছেলি — কিন্তু paper-টির ওপর Editorial Expression of Concern আছে

Personalized generative-AI dialogue experiment conspiracy belief immediateভাবে কমাৰ এবং কয়কে মাস follow-up-এ কিছু effect থাকার report দয়িছেলি। Result suggest করে tailored evidence কিছু strong believer-কওে move করতে পারে। কিন্তু paper বর্তমানৰে Editorial Expression of Concern-এর অধীনৰে; এটি retraction নয়, তবু unresolved methodological/transparency concern-এর signal। তাই clean reading: interesting reported effect, replication/clarification pending — settled fact নয়।

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 has attached an Editorial Expression of Concern to this paper while it evaluates data-handling and reproducibility questions. This is not a retraction and not a finding of misconduct; it means the reported result should be read as provisional until the review is resolved.

Editorial Expression of Concern →

তথ্য, শেষে পর্যন্ত—আর গবেষণাপত্রে ওপর একটি আনুষ্টানিক সতর্কচহ্ন

2024 সালের একটি গবেষণা প্রচলিত ধারণার বিপরীত একটি আকর্ষণীয় ফল জানিয়েছিল। কড়ে যে ষড়যন্ত্র তত্ত্ব সত্যবিশ্বাস করে এবং নজিরে ভাষায় যে প্রমাণ দিয়ে, GPT-4 Turbo-কে যদি সেই নরিদষ্টি প্রমাণের জবাব দিয়ে তনি দফা আলাপ করতে দেওয়া হয়, গড়ে বিশ্বাস প্রায় **20% কম**। এই হ্রাস **দুই মাস** পরেও টকি ছিল। JFK হত্যাকাণ্ড, ভনিগ্রহী, Illuminati-এর মতো ষড়যন্ত্র থেকে COVID-19 ও 2020 US নির্বাচন নিয়ে সমসাময়িক বিশ্বাস—বহির্নির্ন শ্রুতিতেই প্রভাব দেখা যায়; শক্তিশালী বা পরিচয়-সংশ্লিষ্ট বিশ্বাসও। পেশাদার তথ্য-যাচাইকারীরা AI-এর নমুনার **128টি তথ্যভিত্তিক দাবি** পরীক্ষা করে 127টিকে সত্য, 1টিকে বিভিন্নান্তকির এবং 0টিকে ভুল বলছেন।

ছড়িয়ে পড়া শরিরে নাম ছিল: খরগোশের গর্ত থেকে তথ্য দিয়ে মানুষকে ফরিয়ে আনা যায়; ষড়যন্ত্র বিশ্বাসী প্রমাণের বাইরে নয়, বরং সাধারণ খণ্ডনের বদলে তাদের **নজিস্ব উদ্ভূত প্রমাণ** নরিদষ্টিভাবে উত্তর করতে হবে। দাবি সত্যহি আকর্ষণীয়, এবং পরে চনা গবেষণার তুলনায় নকশা সতর্ক।

তারপর জুন 2026-এ *Science* গবেষণাপত্রে ওপর **Editorial Expression of Concern** দিয়ে। এই অংশ বাদ দিলে বর্তমান প্রমাণ অবস্থা ভুল বলা হয়।

Expression of Concern কী—আর কী নয়

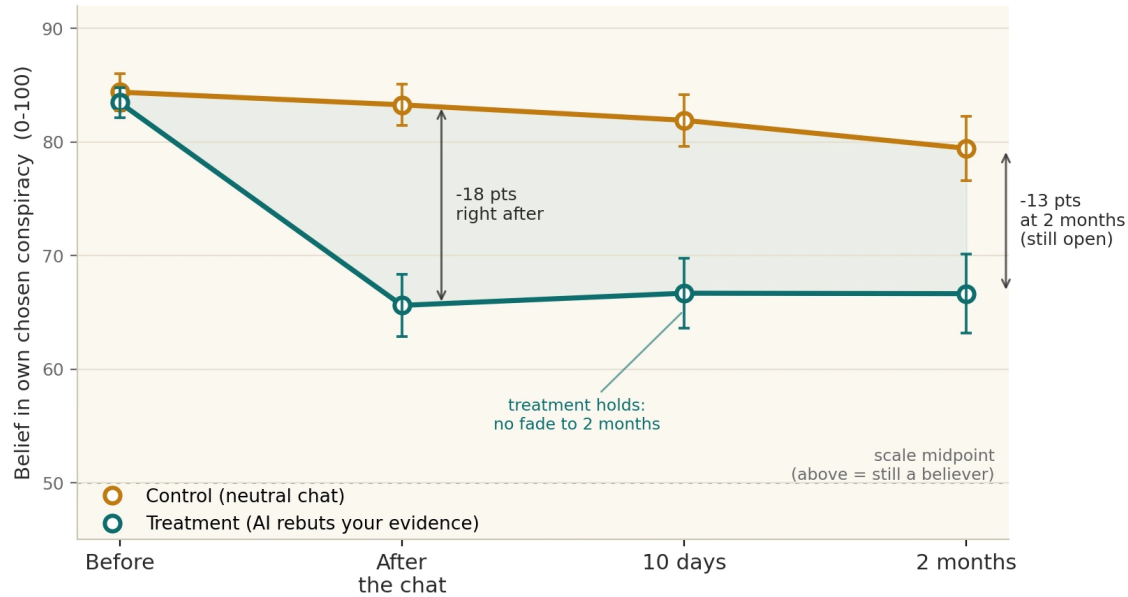
সম্পাদকীয় Expression of Concern সাময়িকীর আনুষ্টানিক সতর্কবার্তা: গবেষণাপত্র নিয়ে প্রশ্ন উঠছে এবং মূল্যায়ন চলছে। এটি **প্রত্যাহার নয়**—গবেষণাপত্র প্রত্যাহৃত নয়; অসদাচরণ সিদ্ধান্তও নজি থেকে নয়। অর্থ: সতর্কতা মাথায় রেখে পড়ুন, নথি বদলাতে পারে। এখানে সমস্যা জানার পরে লেখকেরা তদন্ত করে *Science*-কে নরিদষ্টি সমস্যা ও সংশোধিত বিশ্লেষণ দিয়েছেন।

সতর্কবার্তার কারণটি নরিদষ্টি। অংশগ্রহণকারী বাছাইয়ের মানদণ্ড পাণ্ডুলিপি ও প্রকাশিত বিশ্লেষণ কড়ে একভাবে প্রয়োগ করা হয়নি; জনসাধারণের ডটোস্টে কড়ে ত্রুটির কারণে অতিরিক্ত ও নকল সারিটুক পড়েছিল, ফলে প্রকাশিত উপকরণ থেকে কিছু সংখ্যা পুনরুৎপাদন করা কঠিন হয়। লেখকেরা সংশোধিত ধাপক্রম ও হালনাগাদ ফল *Science*-এ জমা দিয়েছেন; তাদের মতে প্রভাবে **দকি, পরিসংখ্যানগত তাৎপর্য ও বাস্তব মাত্রা** মূল ফলের মতোই থাকে। সাময়িকী এখনো মূল্যায়ন করছে। চূড়ান্ত সিদ্ধান্ত না হওয়া পর্যন্ত নরিদষ্টি সংখ্যাগুলকে অস্থায়ী ধরে নেওয়াই ঠিক।

এটি তাই দুই গল্প: তথ্য গভীরভাবে গুঁথে থাকা বিশ্বাস বদলাতে পারে কনি—একটি সামাজিক-বজ্ঞান ফল; এবং বজ্ঞানিক নথি ত্রুটি পলে প্রকাশ্যে কীভাবে সংশোধন করা হয়—একটি চলমান উদাহরণ।

Belief in a chosen conspiracy, before and after an AI conversation

Study 1: mean belief among the 763 participants who believed their conspiracy, by condition



Reconstructed by The Clean Paper from the public OSF dataset (Costello, Pennycook & Rand, Science 385, eadq1814, 2024). Dataset under a June 2026 Editorial Expression of Concern; values are provisional, not exact published figures. Markers = group means; whiskers = 95% CI; shaded band = treatment-control gap.

প্রতবিদেতি গবেষণায় ব্যক্তিকিনেদরকি তিনি দফার AI খণ্ডন নরিবাচতি ষড়যন্ত্র বশি়াস গড় ~20% কমিয়েছেলি; অসংশ্লিষ্ট নয়িন্তরণ আলাপ তা করনো। পার্থক্য 10 দিন ও 2 মাসেও পরমাপ্যে গ্য ছলি, কনিত্ত আংশকি—অধিকাংশ অংশগ্রহণকারী এখনও ষড়যন্ত্র কছিত্তা বশি়াস করছেলি। জুন 2026 থকে গবেষণাপত্র *Science* সম্পাদকীয় Expression of Concern-এর অধিনে; সংশোধিত বশি়াষণ সাময়িকী মূল্যায়ন করছে, তাই চার্টের মানগুলো অস্থায়ী।

Original figure — The Clean Paper CC BY 4.0

গবেষকরো কী করছেন

- দুই পরীক্ষায় মোট **2,190 অংশগ্রহণকারী** নজিদেরে ভাষায় একটি বশি়াস করা ষড়যন্ত্র তত্ত্ব এবং সমর্থনকারী প্রমাণ লেখেন।
- প্রত্যেকে GPT-4 Turbo-র সঙ্গে **তিনি দফার আলাপ** করেন। চকিৎসায় AI অংশগ্রহণকারীর নরিদৃষ্টি প্রমাণ খণ্ডন করে; নয়িন্তরণে অসংশ্লিষ্ট বিষয় আলাে চনা করে।
- নরিবাচতি ষড়যন্ত্র বশি়াস আলাপের আগে/পরে, তারপর **10 দিন ও 2 মাস** অনুসরণকালে পরমাপ করা হয়।
- নমুনার AI আউটপুটেরে সব **128 তথ্যভিত্তিক দাবি** পশোদার তথ্য-যাচাইকারী মূল্যায়ন করেন।
- প্রভাব অসংশ্লিষ্ট ষড়যন্ত্রে ছড়িয়ে পড়ে কনা এবং **সত্যহিওয়া ষড়যন্ত্র**-তেও সার্বকি সন্দেহপ্রবণতা তরৈকিরে কনা নরিদৃষ্টিতা পরীক্ষা করা হয়।

তাঁরা কী পয়েছেন

- চকিৎসা বনাম নয়িন্তরণে নরিবাচতি ষড়যন্ত্র বশি়াস গড় প্রায় **20% কম**। গবেষণা 1-এ 100-বিন্দু পরসিরে প্রতবিদেতি প্রভাব 95% CI [13.8, 19.7] বিন্দু, $P < 0.001$ ।

- **প্রভাব টকি ছিল।** চকিৎসা দলরে হ্রাসপ্রাপ্ত বিশ্বাস 10 দিন ও 2 মাস অনুসরণকালে আলাপের ঠিকি-পররে স্তররে কাছাকাছি ছিল; নযিন্তরণ উচ্চতর। লখেকরো “অন্তত দুই মাস অকম্বণন” ভাষা ব্যবহার করনে।
- ভনিন ষড়যন্তর ধরন ও শক্তশালী প্রাথমকি বিশ্বাসীতওে প্রভাব দকি থাকে।
- AI দাবিনরিভুলতা নমুনায 127/128 (99.2%) সত্য়, 1/128 (0.8%) বভিন্নান্তকির, 0 ভুল।
- হস্তকষপে সত্য় ষড়যন্তররে বিশ্বাস কমায়নি—সারবকি নরোশ্যবাদ ন্য, অসমরথতি বিশ্বাস লক্ষ্য করার নরিদযিটতা সমরথন করে।
- অসংশ্লযিট ষড়যন্তর বিশ্বাস সামান্য ~12% কম; অংশগ্রহণকারী ষড়যন্তর দাবি প্রতরিে িধ করার ইচ্ছা বাড়ায়। গবষণে 2 প্রধান প্রভাব প্রতলিপি তরৈকর; নযিন্তরণহীন ছে িটতর নমুনা একই দকি দলিওে প্রমাণ দূর্বলতর।

এটকি প্রমাণ করে না

- গবষণাপত্র এখন প্রশ্নহীন প্রমাণ ন্য। তথ্য-handling/পুনরুৎপাদনযে িগ্যতা সমস্যা আনুষ্টানকি সাময়িকি পর্যালোচনায়; সংশোধতি সুনরিদযিট সংখ্যাগুলে ি স্বাধীনভাবে/সাময়িকি-নশিচতি না হওয়া পর্যন্ত অস্থায়ী।
- AI ষড়যন্তর বিশ্বাস “নরিময়” করে না। ~20% গড় হ্রাস অর্থবহ, মুছে ফলো ন্য—অধিকাংশ অংশগ্রহণকারী পরওে বিশ্বাস রাখে।
- বাস্তব জগতরে ব্যবহারকি মৌতায়নে ফল ন্য। গবষণে অংশগ্রহণকারী স্বচেছায় অংশ নযিে ভালো ি সদচিছায় AI-এর সঙগে আলোচনায় অংশ নযিছেন; বাস্তব জগত অধিকাংশ দৃঢ়বিশ্বাসী বিশ্বাসী খণ্ডন বট ষ্ট্রজব কনা অন্য প্রশ্ন।
- 99.2% নরিভুলতা এই কাজ/মডলে/প্রম্পট-এর নমুনা; ভাষা মডলে সাধারণত সত্য়নযিট—এমন নশিচতি ন্য। একই ব্যক্তকিনেদ্রকি পররে িচনা ভুল বিশ্বাস বাড়াতওে ব্যবহার হত পারে।
- “স্থায়ী” এখান দুই মাস, স্থায়ী ন্য।

প্রমাণ কতটা শক্তশালী?

নকশার শক্তি বাস্তব: চকিৎসা/নযিন্তরণ পরীকষা, অসংশ্লযিট-বযিয প্লাসবিো-শলৌ নযিন্তরণ, দুই-গবষণে প্রতলিপি, স্থায়িত্ব অনুসরণকালীন এবং AI তথ্য নরিভুলতা যাচাই। দকি বভিন্ন পরীকষায় সামঞ্জস্যপূর্ণ।

কনিতু **Expression of Concern** পাদটীকা ন্য। প্রকাশতি কৌড/তথ্য থকে সংখ্যা পুনরুৎপাদন করা পুনরুৎপাদনযে িগ্যতার মূল অংশ, এবং ঠকি এখান বোছাই অসামঞ্জস্য ও একত্রতি সারসিমস্যা হয়ছে। লখেকরো সংশোধতি পাইপলাইন প্রভাব সংরক্ষণ করে বলছেন—আশাব্যঞ্জক, কনিতু এট লখেকদরে ববিরণ; *Science*-এর মূল্যায়ন স্বাধীন ধাপ।

সতর্ক অবস্থা: সতর্কতাচহ্নি; খণ্ডতিও ন্য, নতুন করে নশিচতিও ন্য। চৌখে পড়ার মতো ি, ভালো িভাবে-সুপরকিল্পতি বলে মনে হওয়া গবষণে যার সুনরিদযিট পরিমাণগত নথি আনুষ্টানকি পর্যালোচনায়।

কনে এটি গুরুত্বপূর্ণ

প্রভাবশালী নরোশ্যপূর্ণ দৃষ্টিভিঙগি বলে ষড়যন্তর বিশ্বাস অভদে/মনস্তাত্ত্বকি প্রয়োজন-চালতি, তথ্য দযিে তর্ক করে বদলানো যায় না। ফল টকি থাকলে স্থায়ী ব্যক্তকিনেদ্রকি প্রমাণ প্রতকিরিয়া সহৈ নরোশ্যকে পাল্টা প্রমাণ দযে: সাধারণ খণ্ডন ন্য, বিশ্বাসী য়ে নরিদযিট প্রমাণ উদ্ধৃত করে AI বহু পরসিরে তা উত্তর দতি পারে।

গবষণাপত্রটির বর্তমান অবস্থাটিও বজ্জিগ্ননরে একট শকিষা। বহুল প্রশংসতি ফলকে তরুটমুক্ত ধরে নেওয়া বজ্জিগ্ননকি গুণ ন্য; অসামঞ্জস্য ধরা পড়া, তথ্য আবার চালানো ি, সাময়িকি জ্ঞানানো এবং প্রকাশ্য সংশোধন-প্রকরিয়া চালু হওয়াই ব্যবস্থার কাজ করার লক্ষণ—কলেঙ্কারি

নয়। তাই সঠিক অবস্থান অন্ধ প্রশংসা বা তড়িঘড়ি খারজি—কোনো টিহি নয়; অপেক্ষা করা।

AI শিক্ষা দুই দিকের: ব্যক্তিকেন্দ্রিক যুক্তিভুল বিশ্বাস কমাতে পারলে ব্যক্তিকেন্দ্রিক ভুলতথ্য ভুল বিশ্বাস বাড়াতোও পারে। কৌশল নরিপক্ষে; লক্ষ্য নরিপক্ষে নয়।

সংক্ষিপ্ত সারাংশ

2024 গবেষণায় GPT-4 Turbo-র তনি-দফা ব্যক্তিকেন্দ্রিক আলাপ অংশগ্রহণকারীদের নরিবাচতি ষড়যন্ত্র বিশ্বাস গড় ~20% কমিয়েছে; প্রতিবেদিত প্রভাব 2 মাস টকি ছিল এবং নমুনাকৃত AI তথ্যভিত্তিক দাবিগুলো প্রায় সবক্ষেত্রে নরিভুল ছিল। জুন 2026-এ *Science* গবেষণাপত্রে সম্পাদকীয় Expression of Concern দিয়ে অংশগ্রহণকারী-বাছাই অসামঞ্জস্য ও জনসাধারণের ডটোসটে কৌড একত্রীকরণ ত্রুটি নিয়ে। লেখকরা বলনে সংশোধিত বিশ্লেষণ প্রভাবের দিক, তাৎপর্য ও আকার সংরক্ষণ করে; সাময়িকি এখনও মূল্যায়ন করছে। তাই ফল সত্যিই আকর্ষণীয় এবং সতর্কভাবে সুপরিকল্পিত বলে মনে হওয়া, কিন্তু **বর্তমানে আনুষ্ঠানিক পর্যালোচনার অধীনে**—মীমাংসতি নয়, খণ্ডতিও নয়।

উৎস

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>