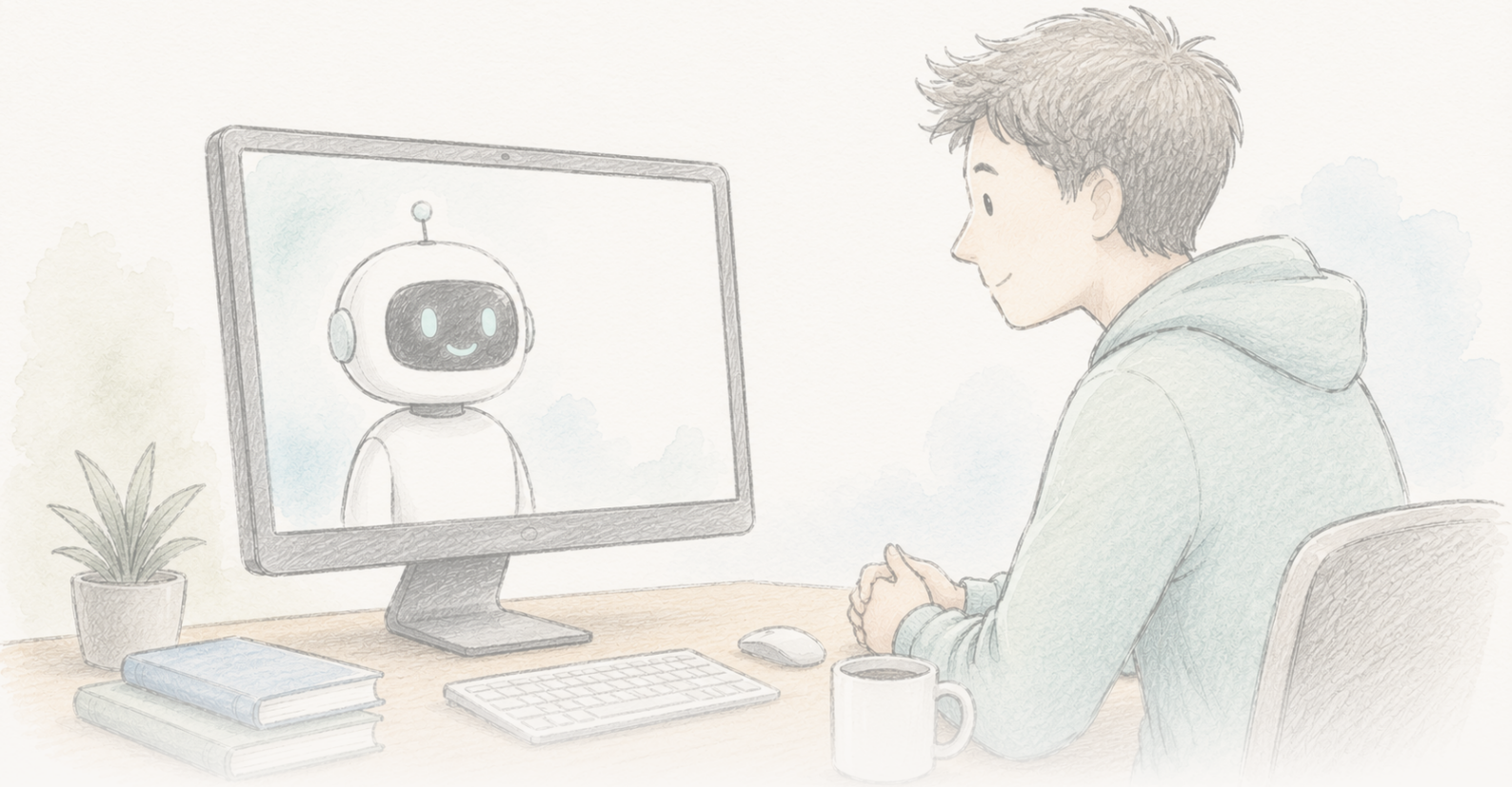


The Clean Paper

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

# Chatbot-এর সঙ্গে কথোপকথন conspiracy belief কয়েক মাস কমিয়ে দিয়েছিল — কিন্তু paper-টির ওপর Editorial Expression of Concern আছে

*Personalized generative-AI dialogue experiment conspiracy belief immediateভাবে কমার এবং কয়েক মাস follow-up-এ কিছু effect থাকার report দিয়েছিল। Result suggest করে tailored evidence কিছু strong believer-কেও move করতে পারে। কিন্তু paper বর্তমানে Editorial Expression of Concern-এর অধীনে; এটি retraction নয়, তবে 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, শেষ পর্যন্ত—আর গবেষণাপত্রের ওপর একটি আনুষ্ঠানিক চিহ্নিতকরা

2024 সালের একটি গবেষণা প্রচলিত ধারণার বিপরীত একটি আকর্ষণীয় ফল জানিয়েছিল। কউে যে ষড়যন্ত্র তত্ত্ব সত্য বিশ্বাস করে এবং নিজের ভাষায় যে প্রমাণ দেয়, GPT-4 Turbo-কে যদি সেই নির্দিষ্ট প্রমাণের জবাব দিয়ে তিনি দফা আলাপ করতে দেওয়া হয়, গড়ে বিশ্বাস প্রায় **20% কম**। এই হ্রাস **দুই মাস** পরও টিকে ছিল। JFK হত্যাকাণ্ড, aliens, Illuminati-এর মতো পুরোনো ষড়যন্ত্র থেকে COVID-19 ও 2020 US election নিয়ে সমসাময়িক বিশ্বাস—বিভিন্ন শ্রণেতিই প্রভাব দখো যায়; শক্তিশালী বা পরচিয়-সংশ্লষ্টি বিশ্বাসওে। পশোদার fact-checker-রা AI-এর নমুনার **128টি তথ্যভত্তিকি দাবি** পরীক্ষা করে 127টিকে সত্য, 1টিকে বভি্রান্তকির এবং 0টিকে ভুল বলছেনে।

Travel করা শরিোনাম ছলি: rabbit গহ্বর থকেে তথ্য দয়িে মানুষকে ফরিয়িে আনা যায়; ষড়যন্ত্র বশি্বাসী প্রমাণরে বাইরে নয়, বরং সাধারণ খণ্ডনরে বদলে তাদরে **নজিস্ব cited প্রমাণ** নর্দিষ্টিভাবে উত্তর করতে হবে। দাবি সত্যহিি আকর্ষণীয়, এবং প্রররোচনা গবেষণার তুলনায় নকশা সতর্ক।

তারপর জুন 2026-এ *Science* গবেষণাপত্ররে ওপর **Editorial Expression of Concern** দয়ে। এই অংশ বাদ দলিে বর্তমান প্রমাণ অবস্থা ভুল বলা হয়।

### Expression of Concern কী—আর কী নয়

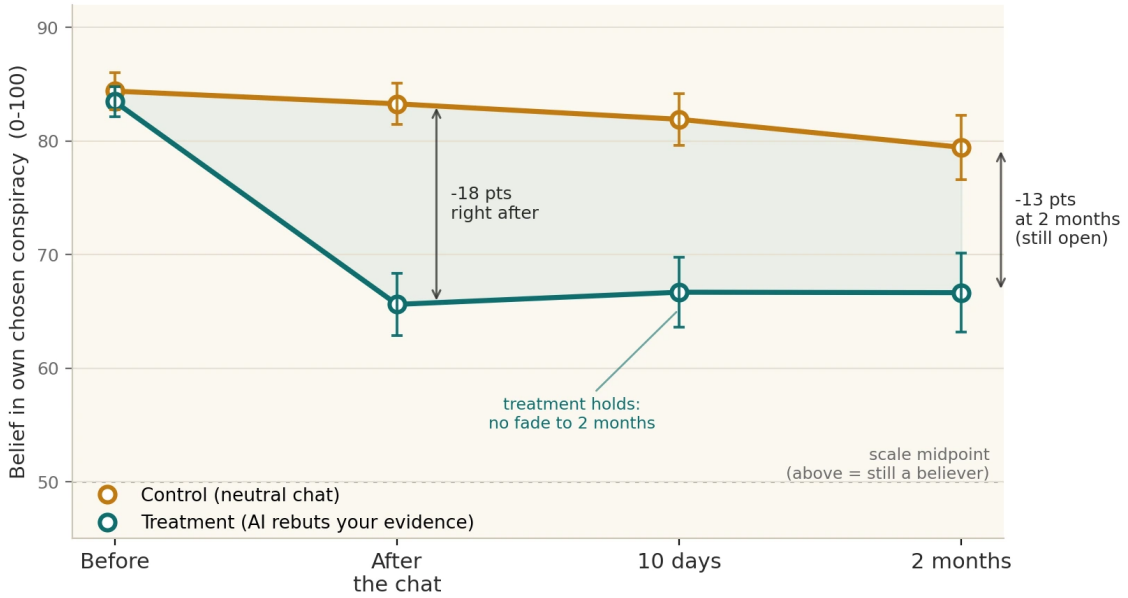
সম্পাদকীয় Expression of Concern সাময়কিীর আনুষ্ঠানকি সতর্কবার্তা: গবেষণাপত্র নয়িে প্রশ্ন উঠছে এবং মূল্যায়ন চলছে। এটি **প্রত্যাহার নয়**—গবেষণাপত্র প্রত্যাহৃত নয়; অসদাচরণ সদ্ধিান্ত-ও নজিে থকেে নয়। অর্থ: সতর্কতা মাথায় রখেে পড়ুন, নথি বদলাতে পার।ে এখানে সমস্যা জানার পরে লখেকেরা তদন্ত করে *Science*-কে নর্দিষ্টি সমস্যা ও সংশোধতি বশি্লষেণ দয়িছেনে।

সতর্কবার্তার কারণটি নর্দিষ্টি। অংশগ্রহণকারী বাছাইয়রে মানদণ্ড পাণ্ডুলপিি ও প্রকাশতি বশি্লষেণ কোডে একভাবে প্রয়োগ করা হয়নি; জনসাধারণরে ডটোসটেে কোডংি ত্রুটরি কারণে অতরিক্তি ও নকল সারি ঢুকে পড়ছেলি, ফলে প্রকাশতি উপকরণ থকেে কছিু সংখ্যা পুনর্বুৎপাদন করা কঠনি হয়। লখেকেরা সংশোধতি pipeline ও হালনাগাদ ফল *Science*-এ জমা দয়িছেনে; তাদরে মতে প্রভাবরে **দকি, পরসিংখ্যানগত তাৎপর্য ও বাস্তব মাত্রা** মূল ফলরে মতোই থাকে। সাময়কিী এখনো মূল্যায়ন করছ।ে চূড়ান্ত সদ্ধিান্ত না হওয়া পর্যন্ত নর্দিষ্টি সংখ্যাগুলোকে অস্থায়ী ধরে নওেয়াই ঠকি।

এটি তাই দুই গল্প: তথ্য গভীরভাবে গঁথে থাকা বশি্বাস বদলাতে পারে কি না—একটি সামাজকি-science ফল; এবং বজৈ্ঞানকি নথি ত্রুটি পলেে জনসাধারণরেভাবে 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 গবষেণায় ব্যক্তকিনেদ্রকি তনি দফার AI খণ্ডন নর্িবাচতি ষড়যন্ত্র বশ্বিাস গড় ~20% কময়িছেলি; অসংশ্লষ্টি নয়িন্ত্রণ আলাপ তা করনেি। পার্থক্য 10 দনি ও 2 মাসওে পরমিাপযে াগ্য ছলি, কন্তিু আংশকি—অধকিাংশ অংশগ্রহণকারী এখনও ষড়যন্ত্র কছিুটা বশ্বিাস করছেলি। জুন 2026 থকেে গবষেণাপত্র *Science* সম্পাদকীয় Expression of Concern-এর অধীনে; সংশ াধতি বশ্লিষেণ সাময়কিী মূল্যায়ন করছ, তাই চার্টরে মানগুলো অস্থায়ী।

Original figure — The Clean Paper CC BY 4.0

## গবষেকরো কী করছেনে

- দুই পরীক্ষায় ম াট **2,190 অংশগ্রহণকারী** নজিদেরে ভাষায় একটি believed ষড়যন্ত্র তত্ত্ব এবং সমর্থনকারী প্রমাণ লখেনে।
- প্রত্যকেে GPT-4 Turbo-র সঙ্গে **তনি দফার আলাপ** করনে। চকিৎিসায় AI অংশগ্রহণকারীর নর্িদষ্টি প্রমাণ rebut কর; নয়িন্ত্রণে অসংশ্লষ্টি বষিয় আল াচনা কর।
- নর্িবাচতি ষড়যন্ত্র বশ্বিাস আলাপরে আগ/পর, তারপর **10 দনি ও 2 মাস** অনুসরণকালীনে পরমিাপ করা হয়।
- নমুনার AI আউটপুটরে সব **128 তথ্যভত্তিকি দাব** পশোদার তথ্য-যাচাইকারী মূল্যায়ন করনে।
- প্রভাব অসংশ্লষ্টি ষড়যন্ত্র-তে spill over কর কনি া এবং **সত্যহিওয়া ষড়যন্ত্র**-তওে সার্বকি skepticism তরৈি কর কনি া নর্িদষ্টিতা পরীক্ষা করা হয়।

## তাঁরা কী পয়েছেনে

- চকিৎিসা বনাম নয়িন্ত্রণে নর্িবাচতি ষড়যন্ত্র বশ্বিাস গড় প্রায় **20% কম**। গবষেণা 1-এ 100-বন্দিু পরসিরে প্রতবিদেতি প্রভাব 95% CI **[13.8, 19.7] বন্দিুগুল**ো, **P < 0.001**।

- **প্রভাব টকিে ছলি।** চকিৎিসা দলরে হ্রাসপ্রাপ্ত বশি্বাস 10 দনি ও 2 মাস অনুসরণকালীনে just-after-আলাপ স্তররে কাছাকাছি ছলি; নয়িন্ত্রণ উচ্চতর। লখেকরো "অন্তত দুই মাস undiminished" ভাষা ব্যবহার করনে।
- ভন্নি ষড়যন্ত্র ধরন ও শক্তশিালী প্রাথমকি বশি্বাসীতওে প্রভাব দকি থাক।ে
- AI দাবি নরি্ভুলতা নমুনায় 127/128 (99.2%) **সত্য**, 1/128 (0.8%) **বভি্রান্তকির**, 0 **ভুল**।
- হস্তক্ষপে **সত্য ষড়যন্ত্র**-র বশি্বাস কমায়নি—সার্বকি cynicism নয়, অসমর্থতি বশি্বাস লক্ষ্য করার নরি্দষ্টিতা সমর্থন কর।ে
- অসংশ্লষ্টি ষড়যন্ত্র বশি্বাস সামান্য ~12% কম;ে অংশগ্রহণকারী ষড়যন্ত্র দাবি push পছেনে করার ইচ্ছা বাড়ায়। গবষেণা 2 প্রধান প্রভাব প্রতলিপিি তরৈি কর;ে না-নয়িন্ত্রণ ছ োটতর নমুনা একই দকি দলিওে প্রমাণ দুর্বলতর।

## এটি কী প্রমাণ করে না

- গবষেণাপত্র এখন **unquestioned প্রমাণ নয়**। তথ্য-handling/পুনরুৎপাদনযে াগ্যতা সমস্যা আনুষ্ঠানকি সাময়কিী পর্যালে াচনায়; সংশে াধতি সুনরি্দষ্টি সংখ্যাগুল ো স্বাধীনভাব/েসাময়কিী-নশি্চতি না হওয়া পর্যন্ত অস্থায়ী।
- AI ষড়যন্ত্র বশি্বাস "**নরিাময়**" কর ে না। ~20% গড় হ্রাস অর্থবহ, erasure নয়—অধকিাংশ অংশগ্রহণকারী afterward-ও বশি্বাস রাখ।ে
- **বাস্তব জগতরে ব্যবহারকি মে াতায়নে** ফল নয়। গবষেণা অংশগ্রহণকারী opt-in করে ভাল ো সদচ্ছিায় AI-এর সঙ্গে আল োচনায় অংশ নয়েছ;ে বাস্তব জগতে অধকিাংশ দৃঢ়বশি্বাসী বশি্বাসী খণ্ডন বট খুঁজবে কি না অন্য প্রশ্ন।
- 99.2% নরি্ভুলতা **এই কাজ/মডলে/প্রম্পট**-এর নমুনা; ভাষা মডলে সাধারণত truthful—এমন নশি্চয়তা নয়। একই ব্যক্তকিন্দ্রকি প্রর োচনা ভুল বশি্বাস বাড়াতওে ব্যবহার হতে পার।ে
- "স্থায়ী" এখানে **দুই মাস**, permanent নয়।

## প্রমাণ কতটা শক্তশিালী?

নকশার শক্তি বাস্তব: চকিৎিসা/নয়িন্ত্রণ পরীক্ষা, অসংশ্লষ্টি-বষিয় প্লাসবি ো-শলৈী নয়িন্ত্রণ, দুই-গবষেণা প্রতলিপিি, স্থায়ত্বি অনুসরণকালীন এবং AI তথ্য নরি্ভুলতা যাচাই। দকি বভিন্নি পরীক্ষায় সামঞ্জস্যপূর্ণ।

কন্তিু **Expression of Concern footnote নয়**। প্রকাশতি ক োড/তথ্য থকেে সংখ্যা পুনরুৎপাদন করা পুনরুৎপাদনযে াগ্যতার মূল অংশ, এবং ঠকি এখানে বাছাই অসামঞ্জস্য ও একত্রতি সার িসমস্যা হয়ছে।ে লখেকরো সংশে াধতি পাইপলাইন প্রভাব সংরক্ষণ করে বলছেনে—আশাব্যঞ্জক, কন্তিু এটি লখেকদরে ববিরণ; *Science*-এর মূল্যায়ন স্বাধীন ধাপ।

সতর্ক অবস্থা: **সতর্কতাচহ্নিতি, খণ্ডতিও নয়, নতুন করে নশি্চতিও নয়**। চ োখে পড়ার মত ো, ভাল োভাবে-সুপরকিল্পতি বলে মনে হওয়া গবষেণা যার সুনরি্দষ্টি পরমিাণগত নথি আনুষ্ঠানকি পর্যালে াচনায়।

## কনে এটি গুরুত্বপূর্ণ

প্রভাবশালী নরৈাশ্যবাদী দৃষ্টভিঙ্গি বলে ষড়যন্ত্র বশি্বাস অভদে/মনস্তাত্ত্বকি প্রয়ে াজন-চালতি, তথ্য দয়িে তর্ক করে বদলান ো যায় না। ফল টকিে থাকা করলে স্থায়ী ব্যক্তকিন্দ্রকি প্রমাণ প্রতক্রিয়িা সইে pessimism-কে পাল্টা প্রমাণ দয়ে: সাধারণ খণ্ডন নয়, বশি্বাসী যে **নরি্দষ্টি প্রমাণ** উদ্ধৃত করে AI বৃহৎ পরসিরে তা উত্তর দতিে পার।ে

গবষেণাপত্রটরি বর্তমান অবস্থাটওি বজি্ঞানরে একটি শকি্ষা। বহুল প্রশংসতি ফলকে ত্রুটমিুক্ত ধরে নওেয়া বজৈ্ঞানকি গুণ নয়; অসামঞ্জস্য

ধরা পড়া, তথ্য আবার চালানো, সাময়িকীকে জানানো এবং প্রকাশ্য সংশোধন-প্রক্রিয়া চালু হওয়াই ব্যবস্থার কাজ করার লক্ষণ—কলেজ্কোরি নয়। তাই সঠিক অবস্থান অন্ধ প্রশংসা বা তড়িঘড়ি খারিজ—কোনোটিই নয়; **অপেক্ষা করা**।

AI শিক্ষা দুই দিকের: ব্যক্তিনেদ্রকি যুক্তি ভুল বিশ্বাস কমাতে পারলে ব্যক্তিনেদ্রকি misinformation ভুল বিশ্বাস বাড়াতেও পারে। কৌশল নিরপেক্ষ; aim নিরপেক্ষ নয়।

# সংক্ষিপ্ত সারাংশ

2024 গবেষণায় GPT-4 Turbo-র তিন-দফা ব্যক্তিনেদ্রকি আলাপ অংশগ্রহণকারীদের নির্বাচিত ষড়যন্ত্র বিশ্বাস গড় ~20% কমিয়েছিল; প্রতিবেদিত প্রভাব 2 মাস টিকে ছিল এবং sampled AI তথ্যভিত্তিক দাবিগুলো প্রায় সবক্ষেত্রে নির্ভুল ছিল। জুন 2026-এ *Science* গবেষণাপত্রে সম্পাদকীয় Expression of Concern দেয় অংশগ্রহণকারী-বাছাই অসামঞ্জস্য ও জনসাধারণের ডটোসটে কোডকেত্রীকরণ ত্রুটি নিয়ে। লেখকেরা বলেন সংশোধিত বিশ্লেষণ প্রভাবের দিক, তাৎপর্য ও আকার সংরক্ষণ করে; সাময়িকী এখনও মূল্যায়ন করছে। তাই ফল সত্যিই আকর্ষণীয় এবং সতর্কভাবে সুপরিকল্পিত বলে মনে হওয়া, কিন্তু **বর্তমানে আনুষ্ঠানিক পর্যালোচনার অধীনে**—মীমাংসিত নয়, debunked-ও 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