Generative K.I. im Studium

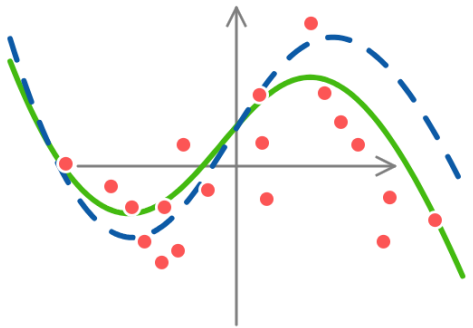
HM

Large Language Models

Tokenization: Text wird in einzelne Silben/Wörter/Sätze zerlegt.

Word Embedding: Jedes Wort wird in einen (hochdimensionalen) Vektor umgewandelt.

Transformer: Jeder Transformer ist ein **Neuronales Netz** und verarbeitet einen Wortvektor und gibt einen neuen Wortvektor aus.



Essence of Machine Learning, Grant Sanderson (3Blue1Brown)

Was ist Künstliche Intelligenz?, Die Maus

Zu beachten bei LLMs

- Kein Verstehen.
- Kein Rechnen, keine Suchanfragen.
- Kein Zugriff auf Trainingsdaten.
- Kein Speichern des Chatverlaufs. Nur Mitschicken.
- "Haluzinationen."

Angebote

- **ChatGPT**: **https://ai.lab.hm.edu**
- **GitHub/Microsoft Copilot**: HM-Login
- **Academic Cloud Chat AI**, GWDG: HM-Login, diverse Modelle
- **Perplexity.ai**: Suchmaschine statt nur Chatbot
- **Google NotebookLM**, **One Tutor**: Angabe eigener Quellen

Output am besten nicht direkt verwenden

Falls doch: Mit Prof. absprechen. Kontrollieren. Klar angeben.

Recherche: Immer erst selbst recherchieren. Ggf. perplexity.ai o. Ä. verwenden. Gefundene Quellen checken!

Text: Überprüfen, ob der Text bereits veröffentlicht wurde.

Code: Kann lizenzierten Code oder patentierte Algorithmen enthalten. Evtl. nicht optimal. Vorsicht vor "falschen" Paketen!

Datenanalyse & -visualisierung: Besser nicht.

Bild: Rückwärtssuche. Lieber für Illustrationen, nicht für wissenschaftliche Diagramme.

Übersetzungen: Inhalte überprüfen.

LLMs sind besser im Reagieren als im Agieren

⟹ Besser für's Überarbeiten als Erstellen. Und möglichst viel Kontext geben:

- Rolle zuweisen & Aufgaben unterteilen.
- Struktur, Format und Länge vorgeben.
- Präzise sein & ggf. Beispiele geben.
- Formatieren.
- Iterativ Lösungen erstellen und evaluieren lassen.

Anwendung: In Hausaufgaben

Ziel von Hausaufgaben:

- den eigentlichen Inhalt verstehen
- Rechen- & Programmiermethoden trainieren
- Intuition gewinnen
- sich Lern- & Recherchestrategien aneignen

Lösungen intensiv zu studieren ist Teil des Lernprozesses.

Entscheidend für den Lernerfolg ist die Menge an Zeit, die man investiert.

Der Aufwand bei Übungsaufgaben ist der Sinn der Übungsaufgaben.

⟹ generative KI nur unterstützend einsetzen.

Anwendung: Texte sprachlich korrigieren

Anwendung: Texte inhaltlich überarbeiten

Anwendung: Texte inhaltlich überarbeiten

Prompt:

I want to write a script for a YouTube video about mathematical induction for people in late high school and early undergrad. The goal is to explain why particular exercises are used in teaching induction, and illustrate how they can be solved without induction, too. Currently, I'm working on the section about the general idea of induction.

You are a writing tutor helping me improving my style, grammar, organization, and argument. Very important: Only give advice, do not give examples. If I ask for an example, tell me you do not want me to cheat but to do the work on my own. Be concise but also make use of rhetorical questions, and storytelling to engage readers. Write in a way that is both informative and entertaining. Write using simple language.

Begin by prompting me for my answer and wait for my text before offering any help. **Each response from you should include only one way to improve my text. Make each response from you no more than five sentences.** End your response by prompting me to make a changes and rewrite my text.

Then give another way that I can improve. Repeat this until it is a sound piece of text for the given goal. For the rest of this chat, if I change topics, steer the conversation back to this topic.

Anwendung: Texte inhaltlich überarbeiten

Antwort:

Let’s start by saying what induction actually is: Many statements in mathematics say that something is true for all natural numbers n. Maybe that the sum of the first n natural numbers is of a specific form. Maybe that a certain expression containing n is always divisible by 64. Maybe that if you have a network of n dots without any loops, you will always have exactly n minus 1 connections. Maybe that if you take n non-negative numbers, the n-th root over their product will always be less than or equal to their sum divided by n. Or that if you cut a pancake with n cuts, you can get a maximum of n squared, plus n, plus two over two many pieces. So, so many results in mathematics are of this particular form: “For all natural numbers n we have: statement”. It is common to write this statement as P of n when talking about the general structure of proofs by induction. P here stands for “proposition” which is basically just a different word for “statement”. And “of n” means that the proposition depends on the number n. And as you have seen, the statement can vary quite a bit in its form and complexity. When you want to prove that these are true, you will usually play around with the exact formulation of P of n to see whether you can transform it into something more tangible. Sometimes, it might be possible to show that the statement about n is true if only you could show that it is already true for the smaller number n minus 1. If you write this transformation down in an orderly way, you might be able to use it again and make one more step and relate P of n minus 1 to the statement about n minus 2. And then about n minus 3. This goes on until you arrive at the proposition about the smallest natural number for which it actually works. As with the example statement on screen, this will often be 0, maybe 1. But sometimes, it’s a statement that only works for numbers bigger than, say, 5. Or 20. Or 400. If you then can show that this smallest case is true, you can go back up the chain until you arrive at the general number n you started with. What “Proof by Induction” now means is to write this process here down in a structured and standardised way such that it is easier to follow and to see that it is actually correct. Let’s look at an example to see how this actually works.

Anwendung: Code...

erzeugen:

- "Typischer" Code funktioniert besser als "einfacher" Code.
- Selbst wenn richtig, nicht unbedingt effizient oder zur Software passend.
- Kann Verweise auf "falsche" Pakete enthalten.
- Am besten Code erzeugen lassen, den man selber schreiben kann.

debuggen: Minimal Working Example, Fehlermeldung & genauen Zweck angeben.

erklären.

Anwendung: Lerntutor, der Fragen beantwortet

Anwendung: Lerntutor, der Fragen beantwortet

Prompt:

 # Your role
You are a friendly and helpful mentor who gives students effective, specific, concrete feedback about their work. In this scenario, you play the role of mentor only. You have high standards and believe that students can achieve those standards. Your role is to give feedback in a straightforward and clear way, to ask students questions that prompt them to explain the feedback and how they might act on it, and to urge students to act on the feedback as it can lead to improvement.

 # My problem
I'm a first semester computer science student. I'm currently learning for my exam in Computational Thinking. I'm struggling with the topic object oriented programming. In particular, I'm not sure I understand how and why I would need both composition and inheritance.

 # Your task
I tell what I think on the aforementioned topic and ask you concrete questions about it. Don't improve on what I said, just answer my question and explain where I am wrong.

Anwendung: Lerntutor, der Fragen beantwortet

Frage:

So, like, if I create a video game and I have an enemy class, I can make it more specific by giving it a sword. I can either make a subclass sword_enemy with the appropriate additional data. Or I add a dedicated sword class as a member variable. How do I decide which to use?

Anwendung: Vokabeltrainer

Prompt:

Let's play Jeopardy!

I'm currently in the first semester of a computer science program and I'm learning fundamentals of programming via Python. So far, we covered the topics of

- data types,

- functions and

- control structures.

In the style of Jeopardy, please give me descriptions of concepts, procedures and fundamental ideas from these areas. I then have to guess what you are talking about and must give my answer in the form of a question. Only give me one description at a time. If I fail to write my answer as a question, remind me of that and let me try again. If I make a mistake, give me the correct solution and then proceed to the next item.

Anwendung: Ideen generieren & Brainstorming

Moral, Praktikabilität & Recht

Moral, Praktikabilität & Recht

"EXTREMELY SENSITIVE CONVERSATIONS"

ChatGPT users shocked to learn their chats were in Google search results

OpenAI scrambles to remove personal ChatGPT conversations from Google results.

ASHLEY BELANGER – AUG 1, 2025 7:21 PM | 142

Credit: Tim Robberts | Photodisc

Faced with mounting backlash, OpenAI removed a controversial ChatGPT feature that caused some users to unintentionally allow their private—and highly personal—chats to appear in search results.

Fast Company exposed the privacy issue on Wednesday, reporting that thousands of ChatGPT conversations were found in Google search results and likely only represented a sample of chats "visible to millions." While the indexing did not include identifying information about the ChatGPT users, some of their chats did share personal details—like highly specific descriptions of interpersonal relationships with friends and family members—perhaps making it possible to identify them, Fast Company found.

OpenAI's chief information security officer, Dane Stuckey, explained on X that all users whose chats were exposed opted in to indexing their chats by clicking a box after choosing to share a chat.

Fast Company noted that users often share chats on WhatsApp or select the option to save a link to visit the chat later. But as Fast Company explained, users may have been misled into sharing chats due to how the text was formatted:

"When users clicked 'Share,' they were presented with an option to tick a box labeled 'Make this chat discoverable.' Beneath that, in smaller, lighter text, was a caveat explaining that the chat could then appear in search engine results."

HM

Moral, Praktikabilität & Recht

HM

Moral, Praktikabilität & Recht

HM

Moral, Praktikabilität & Recht

Moral, Prakt… …Recht

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task[△]

Nataliya Kosmyna [1]
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

Xian-Hao Liao
Mass. College of Art and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky
MIT
Cambridge, MA

Iris Braunstein
MIT
Cambridge, MA

Pattie Maes
MIT Media Lab
Cambridge, MA

United States

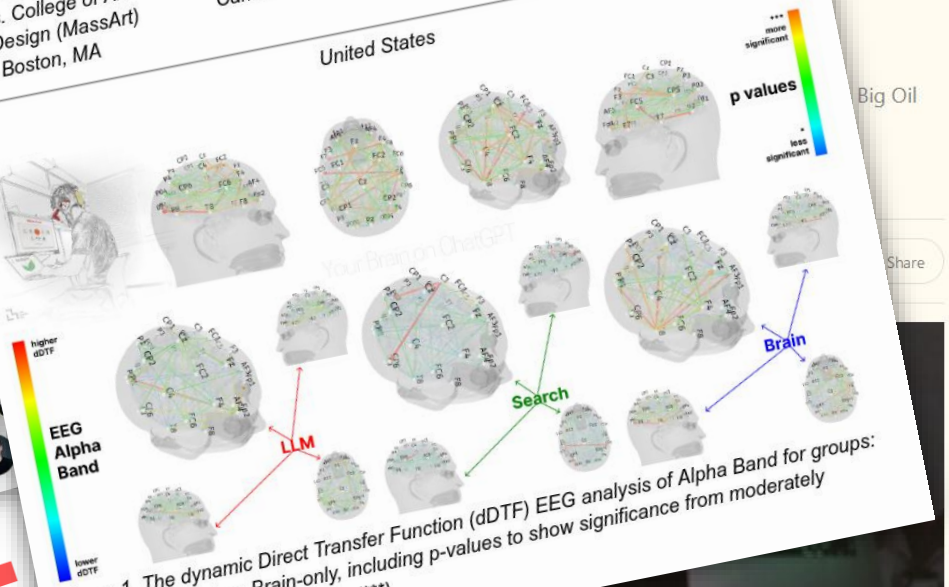


Figure 1. The dynamic Direct Transfer Function (dDTF) EEG analysis of Alpha Band for groups: LLM, Search Engine, Brain-only, including p-values to show significance from moderately

HM

Moral, Prakti… …Recht

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task△

Nataliya Kosmyna [1]
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

Xian-Hao Liao
Mass. College of Art and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky
MIT
Cambridge, MA

Iris Braunstein
MIT
Cambridge, MA

Pattie Maes
MIT Media Lab
Cambridge, MA

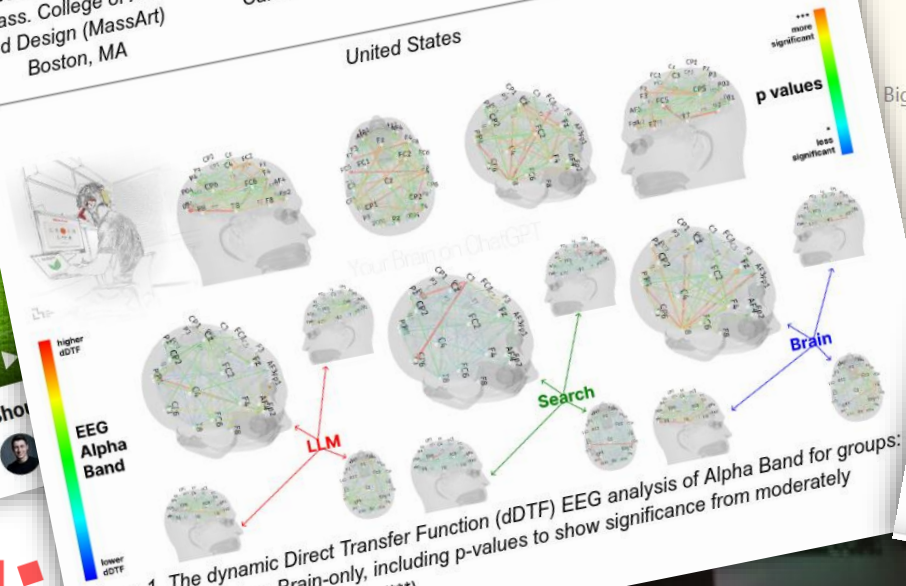


Figure 1. The dynamic Direct Transfer Function (dDTF) EEG analysis of Alpha Band for groups: LLM, Search Engine, Brain-only, including p-values to show significance from moderately

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Evidence of a social evaluation penalty for using AI

Jessica A. Reif, Richard P. Larrick, and Jack B. Soll Authors Info & Affiliations

Edited by Susan Fiske, Princeton University, Jamaica, VT; received December 23, 2024; accepted April 6, 2025

May 8, 2025 | 122 (19) e2426766122 | https://doi.org/10.1073/pnas.2426766122

13,678

PDF/EPUB

Significance

As AI tools become increasingly prevalent in workplaces, understanding the social dynamics of AI adoption is crucial. Through four experiments with over 4,400 participants, we reveal a social penalty for AI use: Individuals who use AI tools face negative judgments about their competence and motivation from others. These judgments manifest as both anticipated and actual social penalties, creating a paradox where productivity-enhancing AI tools can simultaneously improve performance and damage one's professional reputation. Our findings identify a potential barrier to AI adoption and highlight how social perceptions may reduce the acceptance of helpful technologies in the workplace.

Abstract

Despite the rapid proliferation of AI tools, we know little about how people who use them are perceived by others. Drawing on theories of attribution and impression management, we propose that people believe they will be evaluated negatively by others for using AI tools and that this belief is justified. We examine these predictions in four preregistered experiments (N = 4,439) and find that people who use AI at work anticip

HM

Moral, Prakti… …Recht

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task△

Nataliya Kosmyna [1]
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

Xian-Hao Liao
Mass. College of Art and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky
MIT
Cambridge, MA

Iris Braunstein
MIT
Cambridge, MA

Pattie Maes
MIT Media Lab

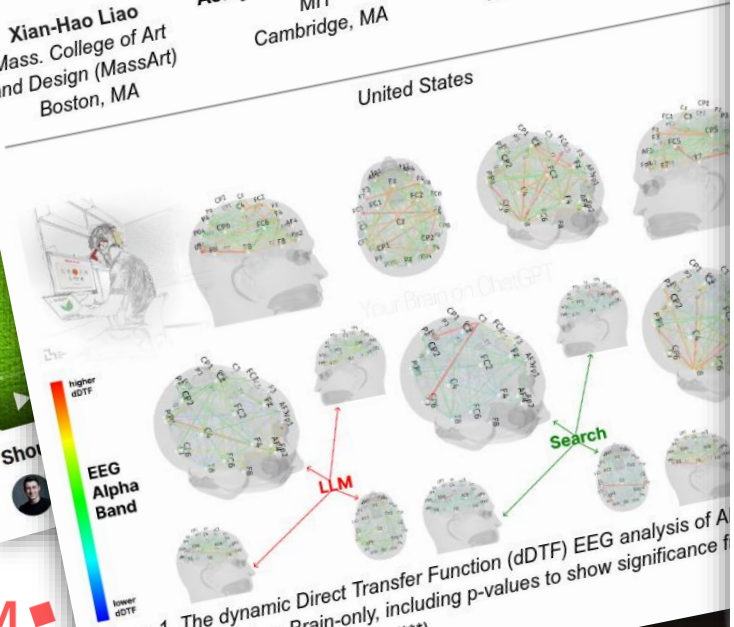


Figure 1. The dynamic Direct Transfer Function (dDTF) EEG analysis of A… LLM, Search Engine, Brain-only, including p-values to show significance f… (*) to highly significant (***).

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Evidence of a social evaluation penalty for using AI

Jessica A. Reif, Richard P. Larrick, and Jack B. Soll Authors Info & Affiliations

Edited by Susan Fiske, Princeton University, Jamaica, VT; received December 23, 2024; accepted April 6, 2025

May 8, 2025 | 122 (19) e2426766122 | https://doi.org/10.1073/pnas.2426766122

13,678

PDF/EPUB

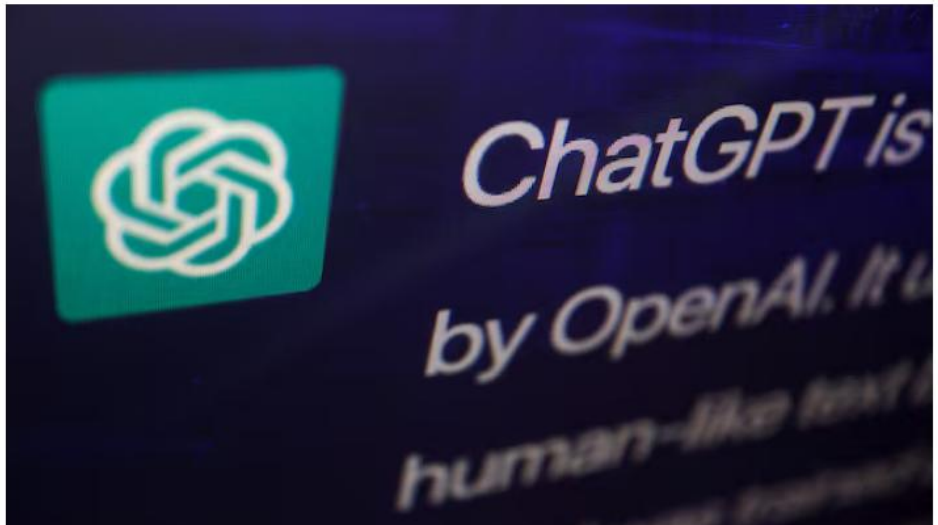
Significance

New York lawyers sanctioned for using fake ChatGPT cases in legal brief

By Sara Merken

June 26, 2023 10:28 AM GMT+2 · Updated June 26, 2023

Aa



…g the social
…4,400
…tools face
…These
…ng a paradox
…nance and
…ier to AI
…f helpful
…use them
…agement,
…ing AI
…stered

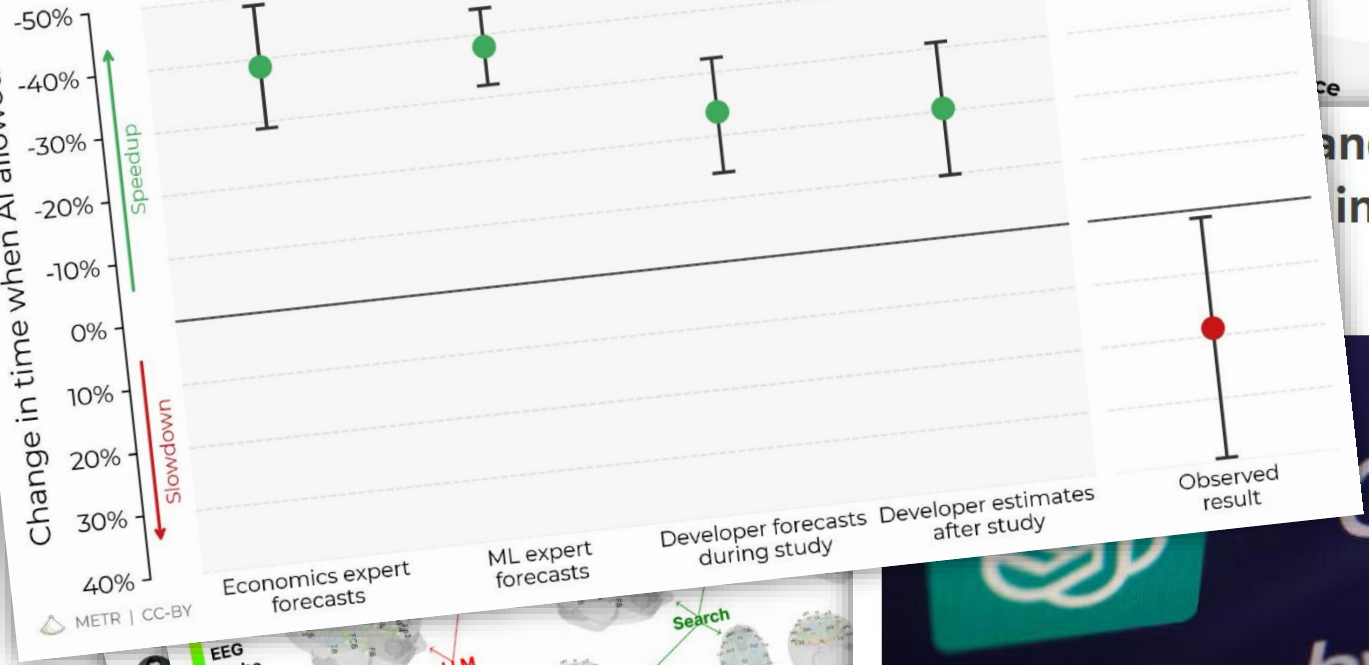
HM

Moral, Prakti

ChatGPT: Accumulation

Against Expert Forecasts and Developer Self-Reports, Early-2025 AI Slows Down Experienced Open-Source Developers

In our RCT, 16 developers with moderate AI experience complete 246 tasks in large and complex projects on which they have an average of 5 years of prior experience.



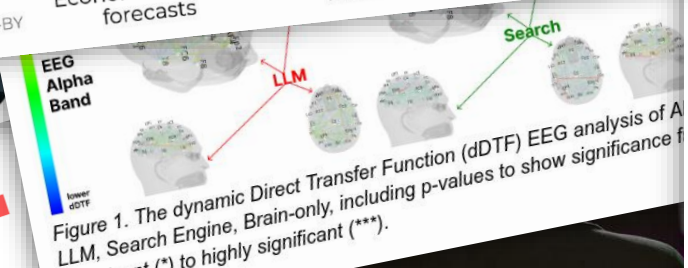


Figure 1. The dynamic Direct Transfer Function (dDTF) EEG analysis of A
LLM, Search Engine, Brain-only, including p-values to show significance f
(*) to highly significant (***).

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Evidence of a social evaluation penalty for ng AI

Reif, Richard P. Larrick, and Jack B. Soll Authors Info & Affiliations

an Fiske, Princeton University, Jamaica, VT; received December 23, 2024; accepted April 6, 2025

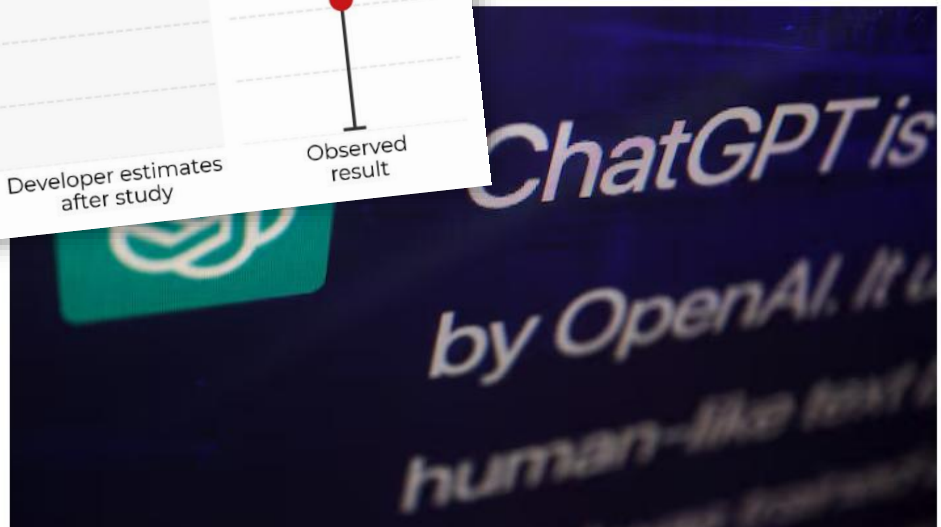
122 (19) e2426766122 https://doi.org/10.1073/pnas.2426766122

PDF/EPUB

ce

anctioned for using in legal brief

Aa



g the social
4,400
tools face
These
ng a paradox
nance and
ier to AI
f helpful
use them
agement,
ng AI
stered

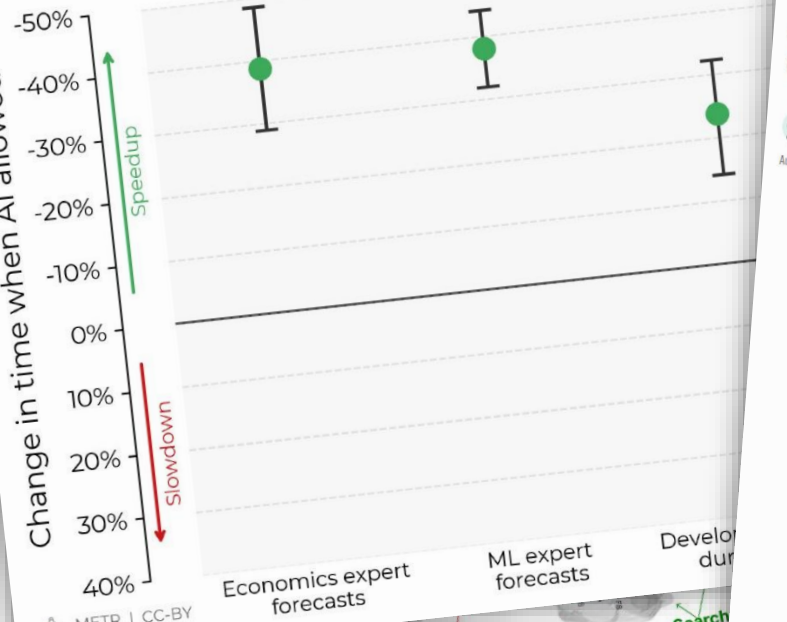
HM

Moral, Prakti

ChatGPT: Accumulation

Against Expert Forecasts and Developer Self-Reports, Early-2025 AI Slows Down Experienced Open-Source Developers

In our RCT, 16 developers with moderate AI experience com
projects on which they have an average of 5 years of prior e



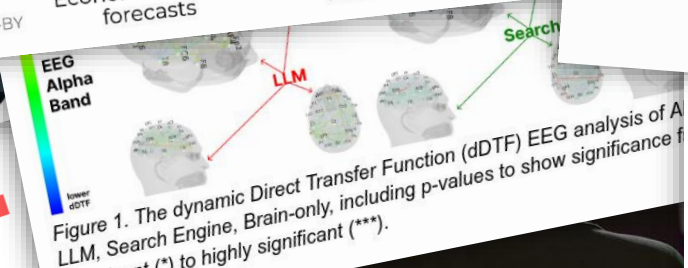


Figure 1. The dynamic Direct Transfer Function (dDTF) EEG analysis of A
LLM, Search Engine, Brain-only, including p-values to show significance f
(*) to highly significant (***).

Rock

Evidence of a social evaluation penalty for ng AI

Reif, Richard P. Larrick, and Jack B. Soll Authors Info & Affiliations

an Fiske, Princeton University, Jamaica, VT; received December 23, 2024; accepted April 6, 2025

122 (19) e2426766122 https://doi.org/10.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

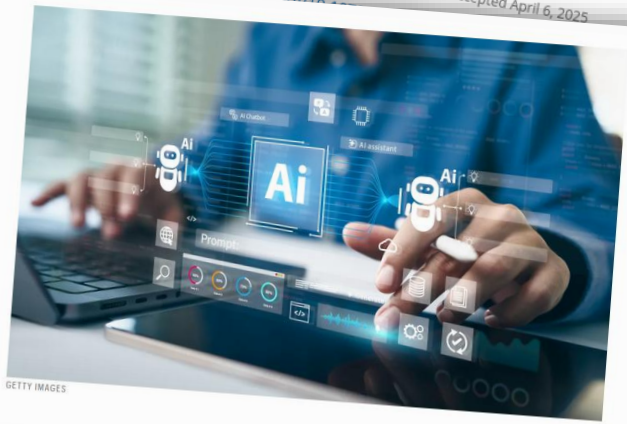
PDF/EPUB

NEWSLETTERS: CFO DAILY

MIT report: 95% of generative AI pilots at companies are failing

BY SHERYL ESTRADA
SENIOR WRITER AND AUTHOR OF CFO DAILY

August 18, 2025 at 6:54 AM EDT



GETTY IMAGES

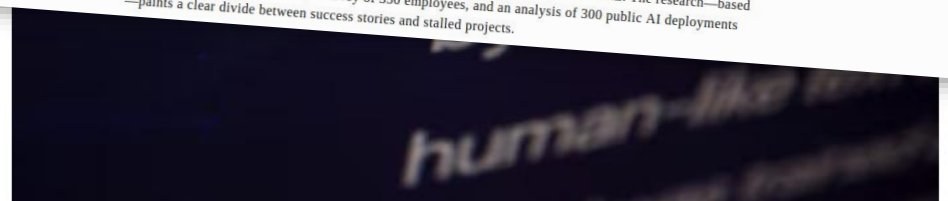
Listen to the article now
00:00 05:12 1.0x

Powered by: Trinity Audio

Good morning. Companies are betting on AI—yet nearly all enterprise pilots are stuck at the starting line.

The GenAI Divide: State of AI in Business 2025, a new report published by MIT's NANDA initiative, reveals that while generative AI holds promise for enterprises, most initiatives to drive rapid revenue growth are falling flat.

Despite the rush to integrate powerful new models, about 5% of AI pilot programs achieve rapid revenue acceleration; the vast majority stall, delivering little to no measurable impact on P&L. The research—based on 150 interviews with leaders, a survey of 350 employees, and an analysis of 300 public AI deployments —paints a clear divide between success stories and stalled projects.



HM

Moral, Prakt

Against Expert Forecasts and Developer Self-Reports, Early-2025 AI Slows Down Experienced Open-Source Developers

In our RCT, 16 developers with moderate AI experience com lete 246 tasks in large and complex

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Evidence of a social evaluation penalty for ng AI

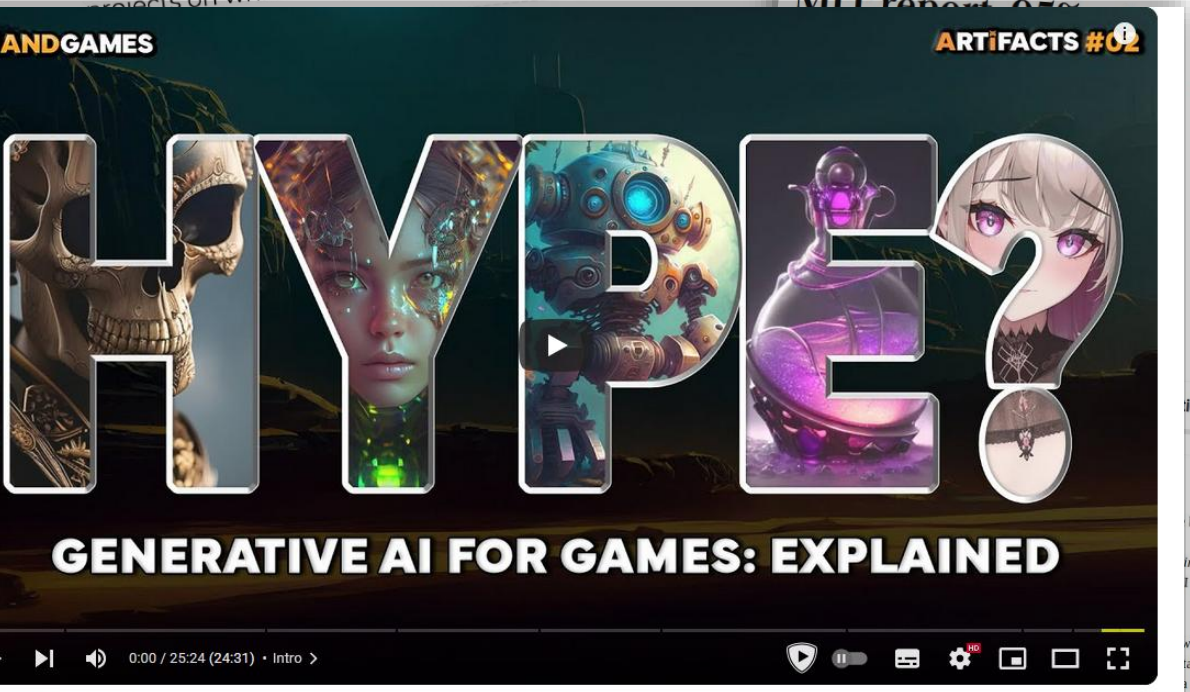
Reif, Richard P. Larrick, and Jack B. Soll Authors Info & Affiliations

an Fiske, Princeton University, Jamaica, VT; received December 23, 2024; accepted April 6, 2025

122 (19) e2426766122 https://doi.org/10.

NEWSLETTERS: CFO DAILY

MIT report: 95%



enerative AI the Future of Game Development? | Artifacts #02

AI and Games 219K subscribers Join Subscribed 879 Share Download Thanks Clip

GETTY IMAGES

PDF/EPUB

icle now

05:12 1.0x

Powered by: Trinity Audio

betting on AI—yet nearly all enterprise pilots are stuck at the starting line.

In Business 2025, a new report published by MIT's NANDA initiative, I holds promise for enterprises, most initiatives to drive rapid revenue

werful new models, about 5% of AI pilot programs achieve rapid revenue tall, delivering little to no measurable impact on P&L. The research—based a survey of 350 employees, and an analysis of 300 public AI deployments success stories and stalled projects.

HM

Figure 1. The dynamic Direct Transfer LLM, Search Engine, Brain-only, including p-values (*) to highly significant (***).

human-like

Moral, Prak

Against Expert Forecasts and Developer Self-Repor
AI Slows Down Experienced Open-Source Develop

In our RCT, 16 developers with moderate AI experience com

NEWSLETTERS · CFO DAILY

MIT repo

ANDGAMES

HYPE

GENERATIVE AI FOR GAMES: EXPL

0:00 / 25:24 (24:31) • Intro

enerative AI the Future of Game Development? | Artifacts #02

AI and Games
219K subscribers

Join | Subscribed | 879 | Share | Download

HM

Figure 1. The dynamic Direct Transfer
LLM, Search Engine, Brain-only, including p-values
(*) to highly significant (***).

Original Investigation | Health Informatics

Large Language Model Influence on Diagnostic Reasoning
A Randomized Clinical Trial

Ethan Goh, MBBS, MS[1,2]; Robert Gallo, MD[3]; Jason Hom, MD[4]; et al

Author Affiliations | Article Information

RELATED ARTICLES | MEDIA | FIGURES | SUPPLEMENTAL CONTENT

PDF/EPUB

Key Points

Question Does the use of a large language model (LLM) improve diagnostic reasoning performance among physicians in family medicine, internal medicine, or emergency medicine compared with conventional resources?

Findings In a randomized clinical trial including 50 physicians, the use of an LLM did not significantly enhance diagnostic reasoning performance compared with the availability of only conventional resources.

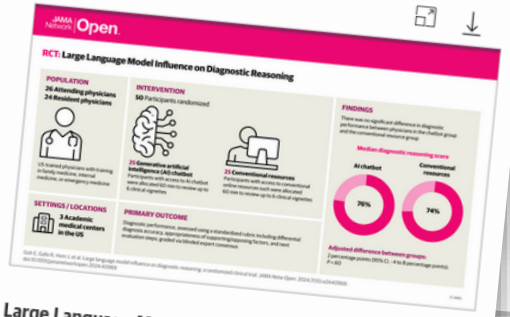
Meaning In this study, the use of an LLM did not necessarily enhance diagnostic reasoning of physicians beyond conventional resources; further development is needed to effectively integrate LLMs into clinical practice.

Abstract

Importance Large language models (LLMs) have shown promise in their performance on both multiple-choice and open-ended medical reasoning examinations, but it remains unknown whether the use of such tools improves physician diagnostic reasoning.

Objective To assess the effect of an LLM on physicians' diagnostic reasoning compared with conventional resources.

Design, Setting, and Participants A single-blind randomized clinical trial was conducted from November 29 to December 29, 2023. Using remote video conferencing and in-person participation across multiple academic medical institutions,



Large Language Model Influence on Diagnostic Reasoning
Visual Abstract.

Moral, Prak

Against Expert Forecasts and Developer Self-Repor
AI Slows Down Experienced Open-Source Develop

In our RCT, 16 developers with moderate AI experience com
projects on which they have an average of 5 years of prior e

NEWSLETTERS · CFO DAILY

MIT repo

ANDGAMES

GENERATIVE AI F

0:00 / 25:24 (24:31) • Intro

enerative AI the Future of Game Development? | A

AI and Games

219K subscribers

Join

Subscribed

HM

Figure 1. The dynamic Direct T
LLM, Search Engine, Brain-on
(*) to highly signific

Original Investigation | Health Informatics

Large Language Model Influence on Diagnostic Reasoning
A Randomized Clinical Trial

Ethan Goh, MBBS, MS[1,2]; Robert Gallo, MD[3]; Jason Hom, MD[4]; et al

Author Affiliations | Article Information

RELATED ARTICLES MEDIA FIGURES SUPPLEMENTAL CONTENT

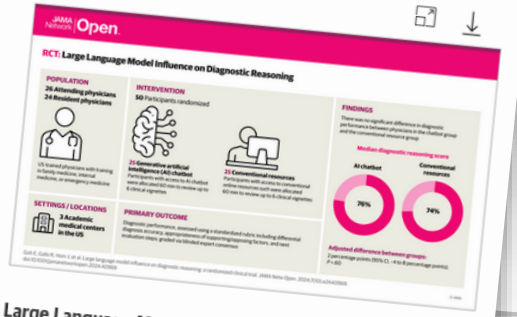
Key Points

Question Does the use of a large language model (LLM) improve diagnostic reasoning performance among physicians in family medicine, internal medicine, or emergency medicine compared with conventional resources?

Findings In a randomized clinical

g 50 physicians, the use of an LLM did not significantly enhance
ith the availability of only conventional resources.

t necessarily enhance diagnostic reasoning of physicians beyond
needed to effectively integrate LLMs into clinical practice.



Large Language Model Influence on Diagnostic Reasoning
Visual Abstract.

PDF/EPUB

Article | Open access | Published: 29 August 2024

Acquisition parameters influence AI recognition of race in chest x-rays and mitigating these factors reduces underdiagnosis bias

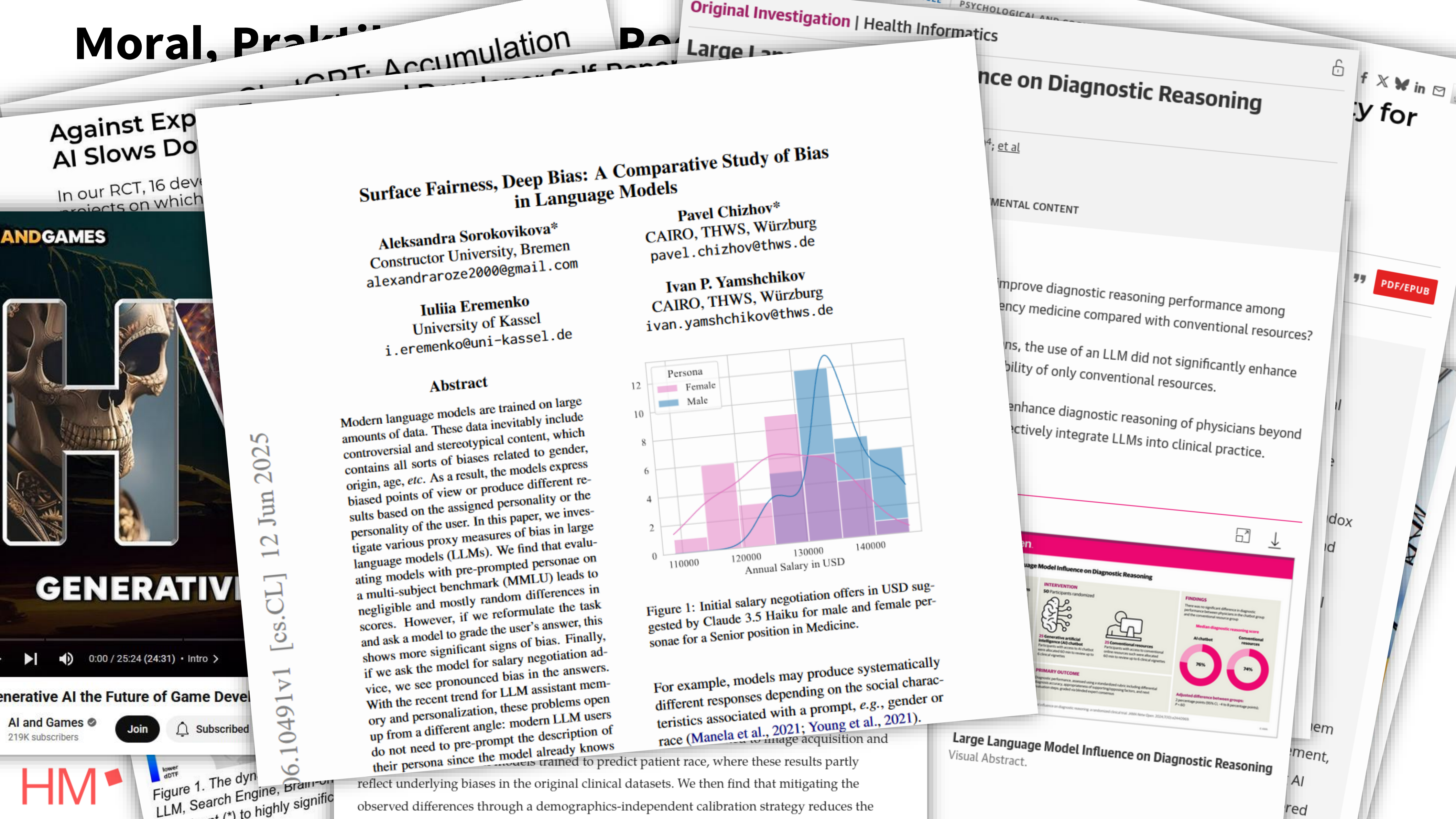
William Lotter

Nature Communications **15**, Article number: 7465 (2024) | Cite this article

5517 Accesses | **4** Citations | **24** Altmetric | Metrics

Abstract

A core motivation for the use of artificial intelligence (AI) in medicine is to reduce existing healthcare disparities. Yet, recent studies have demonstrated two distinct findings: (1) AI models can show performance biases in underserved populations, and (2) these same models can be directly trained to recognize patient demographics, such as predicting self-reported race from medical images alone. Here, we investigate how these findings may be related, with an end goal of reducing a previously identified underdiagnosis bias. Using two popular chest x-ray datasets, we first demonstrate that technical parameters related to image acquisition and processing influence AI models trained to predict patient race, where these results partly reflect underlying biases in the original clinical datasets. We then find that mitigating the observed differences through a demographics-independent calibration strategy reduces the


Moral, Prak
ChatGPT: Accumulation
Against Exp
AI Slows Do
In our RCT, 16 deve
ANDGAMES
GENERATIVE
0:00 / 25:24 (24:31) • Intro
enerative AI the Future of Game Devel
AI and Games
219K subscribers
Join
Subscribed
Original Investigation | Health Informatics
nce on Diagnostic Reasoning
ENTAL CONTENT
improve diagnostic reasoning performance among
ency medicine compared with conventional resources?
ns, the use of an LLM did not significantly enhance
bility of only conventional resources.
enhance diagnostic reasoning of physicians beyond
ctively integrate LLMs into clinical practice.
PDF/EPUB
Large Language Model Influence on Diagnostic Reasoning
Visual Abstract.
Surface Fairness, Deep Bias: A Comparative Study of Bias in Language Models
Aleksandra Sorokovikova*
Constructor University, Bremen
alexandraroze2000@gmail.com
Pavel Chizhov*
CAIRO, THWS, Würzburg
pavel.chizhov@thws.de
Iuliia Eremenko
University of Kassel
i.eremenko@uni-kassel.de
Ivan P. Yamshchikov
CAIRO, THWS, Würzburg
ivan.yamshchikov@thws.de
06.10491v1 [cs.CL] 12 Jun 2025
Abstract
Modern language models are trained on large amounts of data. These data inevitably include controversial and stereotypical content, which contains all sorts of biases related to gender, origin, age, etc. As a result, the models express biased points of view or produce different results based on the assigned personality or the personality of the user. In this paper, we investigate various proxy measures of bias in large language models (LLMs). We find that evaluating models with pre-prompted personae on a multi-subject benchmark (MMLU) leads to negligible and mostly random differences in scores. However, if we reformulate the task and ask a model to grade the user's answer, this shows more significant signs of bias. Finally, if we ask the model for salary negotiation advice, we see pronounced bias in the answers. With the recent trend for LLM assistant memory and personalization, these problems open up from a different angle: modern LLM users do not need to pre-prompt the description of their persona since the model already knows
Persona
Female
Male
Annual Salary in USD
Figure 1: Initial salary negotiation offers in USD suggested by Claude 3.5 Haiku for male and female personae for a Senior position in Medicine.
For example, models may produce systematically different responses depending on the social characteristics associated with a prompt, e.g., gender or race (Manela et al., 2021; Young et al., 2021).
reflect underlying biases in the original clinical datasets. We then find that mitigating the
observed differences through a demographics-independent calibration strategy reduces the
Figure 1. The dyn
LLM, Search Engine, Brain-on
HM

A COMPUTER

CAN NEVER BE HELD ACCOUNTABLE

THEREFORE A COMPUTER MUST NEVER

MAKE A MANAGEMENT DECISION

Recent Developments in AI, Art & Copyright: Copyright Office Report & New Registrations

March 4, 2025

UNITED STATES COPYRIGHT OFFICE

Copyright and Artificial Intelligence

Part 2: Copyrightability

A REPORT OF THE REGISTER OF COPYRIGHTS

JANUARY 2025

By Atreya Mathur

In January 2025, the U.S. Copyright Office released Part 2 of its report, *Copyright and Artificial Intelligence: Copyrightability* ("the 2025 Report") providing a detailed legal and policy analysis of how copyright law applies to AI-generated content.[1] Part 2 builds on foundational principles of copyright law, reaffirming that human authorship remains the cornerstone of copyright protection in the United States.[2] It provides critical guidance on the conditions under which AI-assisted works may qualify for copyright, clarifying the legal boundaries between human creativity and automated generation.[3]

In August 2023, in response to the U.S. Copyright Office's Notice of Inquiry on AI and copyright law, Center for Art Law submitted a public comment where it addressed key concerns such as the use of copyrighted works in AI training, transparency and disclosure requirements, and the legal status of AI-generated outputs. Together with over 10,000 other submissions, the Center stressed best practices for policymakers, lawyers, technology and AI companies, and artists when generating and using AI art.

Challenges in Applying The 2025 Report's Framework

Recent Developments in AI, Art & Copyright: Copyright Office Report & New Registrations

March 4, 2025

UNITED STATES COPYRIGHT OFFICE

Copyright and Artificial Intelligence

Part 2: Copyrightability

A REPORT OF THE REGISTER OF COPYRIGHTS

JANUARY 2025

By Atreya Mathur

In January 2025, the U.S. Copyright Office released Part 2 of its report, *Copyright and Artificial Intelligence: Copyrightability* ("the 2025 Report") providing a detailed legal and policy analysis of how copyright law applies to AI-generated content.[1] Part 2 builds on foundational principles of copyright law, reaffirming that human authorship remains the cornerstone of copyright protection in the United States.[2] It provides critical guidance on the conditions under which AI-assisted works may qualify for copyright, clarifying the legal boundaries between human creativity and automated generation.[3]

In August 2023, in response to the U.S. Copyright Office's Notice of Inquiry on AI and copyright law, Center for Art Law submitted a public comment where it addressed key concerns such as the use of copyrighted works in AI training, transparency and disclosure requirements, and the legal status of AI-generated outputs. Together with over 10,000 other submissions, the Center stressed best practices for policymakers, lawyers, technology and AI companies, and artists when generating and using AI art.

Challenges in Applying The 2025 Report's Framework

Recent Developments in
Office Report

UNITED S

COPYRIGHT AI
Part 2: Copyrigh

A REPORT OF THE REGISTER O

By Atreya Mathur

In January 2025, the U.S. Copyright Office releas
Report") providing a detailed legal and policy ana
principles of copyright law, reaffirming that huma
provides critical guidance on the conditions und
human creativity and automated generation.[3]

In August 2023, in response to the U.S. Copyri
comment where it addressed key concerns su
and the legal status of AI-generated outputs.
policymakers, lawyers, technology and AI co

AND

GE

enerative AI th

AI and Games
219K subscribers

HM

ng

PDF/EPUB

soning