

Τεχνητή Νοημοσύνη: Κίνδυνοι και Επιπτώσεις

Μάριος Δικαιάκος

Τμήμα Πληροφορικής

mdd@ucy.ac.cy





Ισπανική Σχολή Ιππασίας, Βιέννη

Διάβρωση Ακεραιότητας



COLLEGE STUDENTS HAVE ALREADY CHANGED FOREVER

Members of the class of 2026 have had access to AI since they were freshmen. Almost all of them are using it to do their work.

BY IAN BOGOST



The Atlantic • 14h •

A majority of undergraduate students use AI for classwork, several studies show. Ian Bogost spoke with college students to explore their motivations behind using the tool—and the painful revelation professors may face this in classrooms this fall.
<https://theatlantic.com/DESh4BQ>

M. D. Dikaiakos - <http://www.cs.ucy.ac.cy/mdd>

Αποτίμηση Μεγέθους;



Η πλατφόρμα Chegg

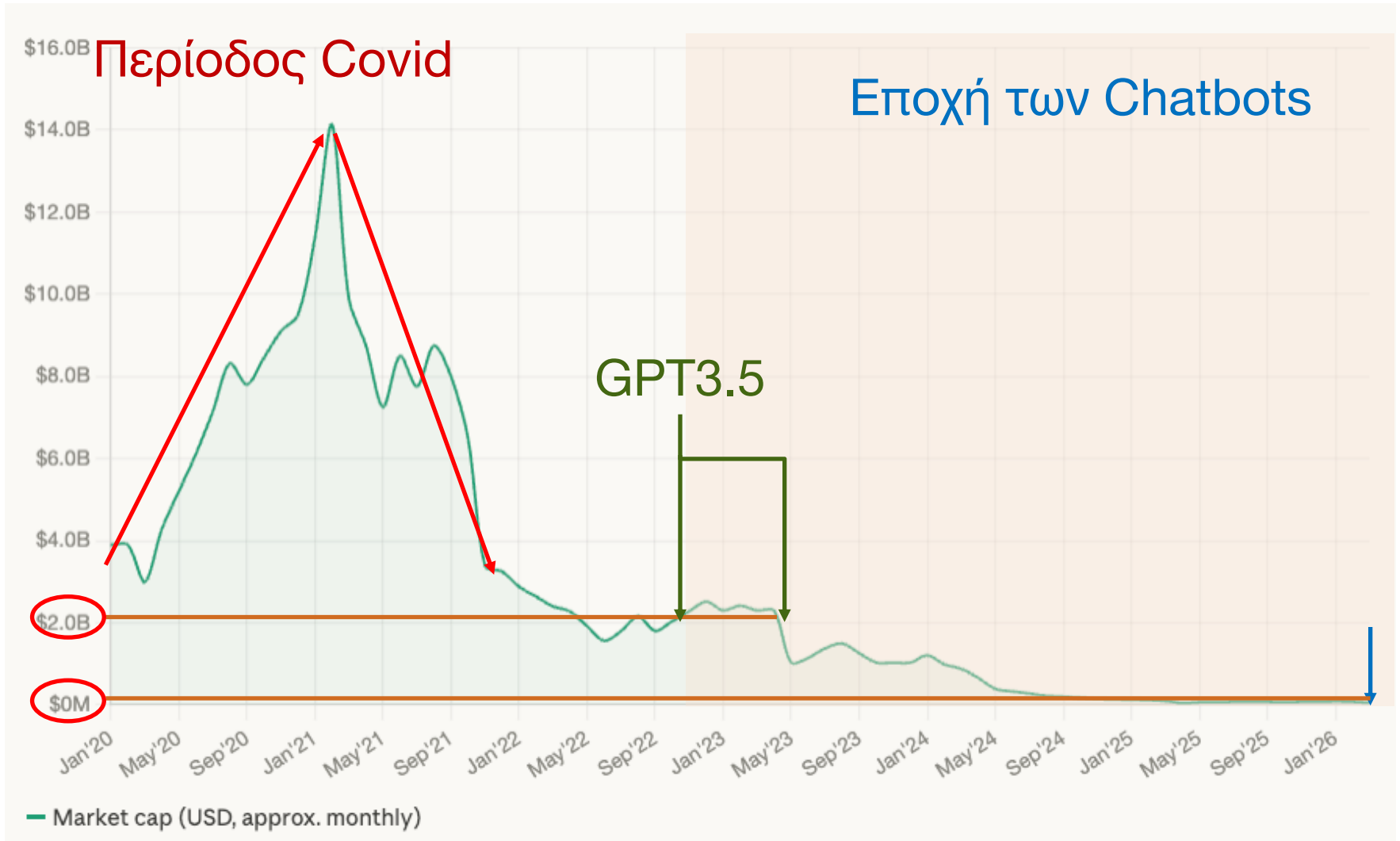
Chegg Chegg
Educational support services company



Chegg, Inc., is an American educational technology company based in Santa Clara, California. It provides **homework help**, digital and physical textbook rentals, textbooks, online **tutoring**, and **other student services**, powered by artificial intelligence. The company has 6.6 million subscribers.

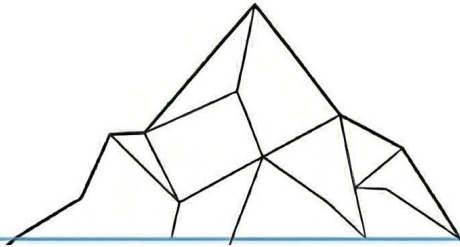
Source: [Wikipedia](#)





Χρηματιστηριακή αξία της Chegg: \$2,3B (4/23) -> \$60M (3/26)

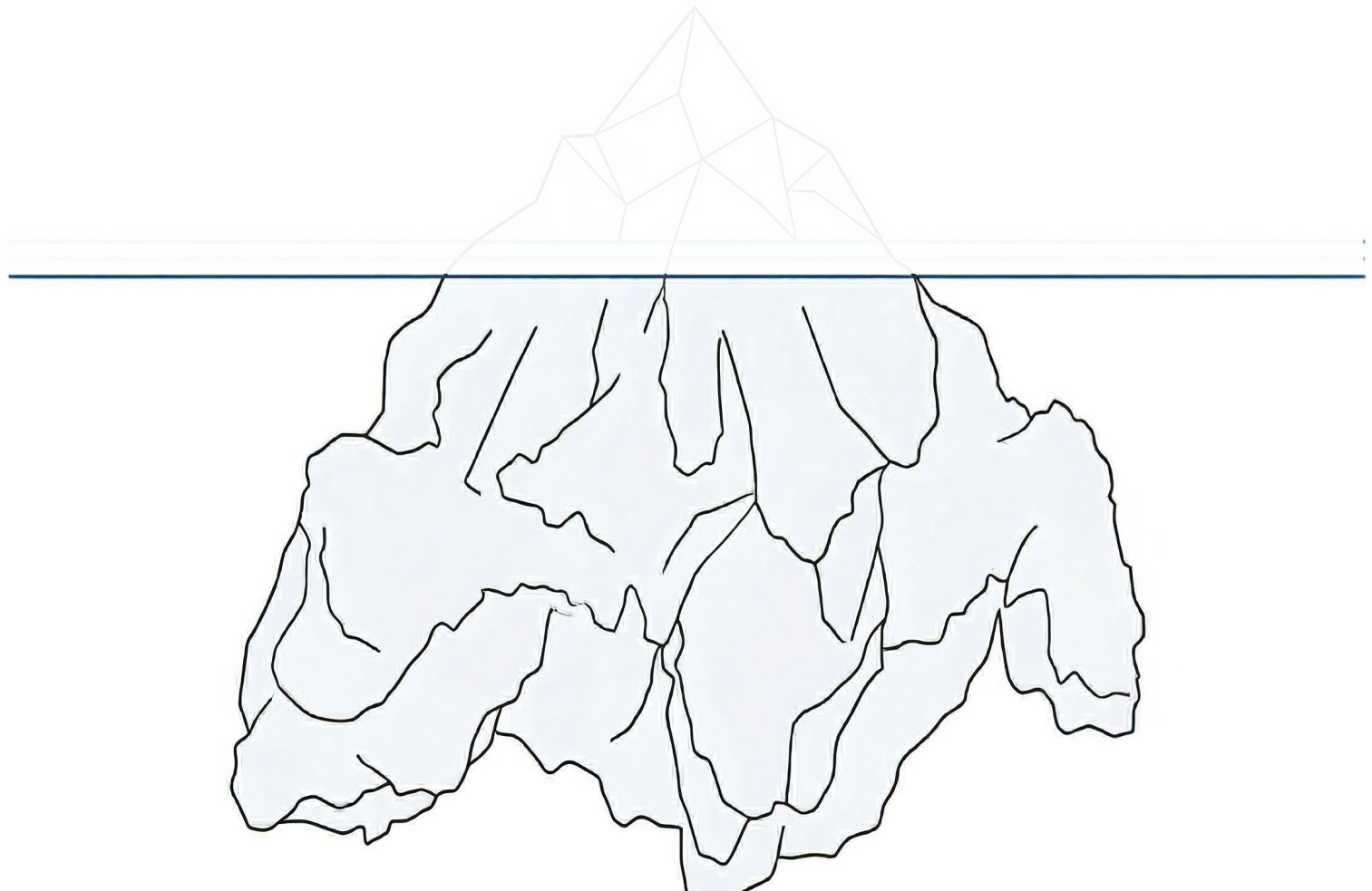
Γνωστική Αυταπάτη



Όταν «καταναλώνουμε» πληροφορίες **χωρίς προσπάθεια** αναπτύσσουμε μια **ψευδαίσθηση μάθησης**.

Μόνο η **ενεργητική ανάκληση** και ο συνειδητός κόπος στην **αναζήτηση**, στην **επανάληψη**, στην **αντιπαραβολή** και στη **σύνθεση** πληροφοριών μπορεί να μετατρέψουν την πληροφορία σε **ανθεκτική μνήμη** και **ουσιαστική γνώση**.

Νοητική Ατροφία



Νοητική Ατροφία



Η μελέτη και η συγγραφή συνιστούν **δομημένες νοητικές διαδικασίες** που βοηθούν τους ανθρώπους να **εμπεδώσουν, οργανώσουν και εξελίξουν** τις ιδέες, τις σκέψεις και τις γνώσεις τους.

Η **συγγραφή είναι από τους πιο επιδραστικούς τρόπους σκέψης.**

Αυτή η ευκαιρία χάνεται όταν η συγγραφή ανατίθεται εξ ολοκλήρου στην ΤΝ.

Απώτερη συνέπεια είναι η **γνωστική ατροφία**: αποδυναμώνεται η ικανότητα του ανθρώπου να απομνημονεύει γνώσεις και πληροφορίες στις οποίες να μπορεί να βασίσει **σύνθετους λογικούς συλλογισμούς, αυτόνομη, κριτική και δημιουργική σκέψη.**

Editorial

<https://doi.org/10.1038/44222-025-00323-4>

Writing is thinking

Check for updates

On the value of human-generated scientific writing in the age of large-language models.

Writing scientific articles is an integral part of the scientific method and common practice to communicate research findings. However, writing is not only about reporting results; it also provides a tool to uncover new thoughts and ideas. Writing compels us to think – not in the chaotic, non-linear way our minds typically wander, but in a structured, intentional manner. By writing it down, we can sort years of research, data and analysis into an actual story, thereby identifying our main message and the influence of our work. This is not merely a philosophical observation; it is backed by scientific evidence. For example, handwriting can lead to widespread brain connectivity and has positive effects on learning and memory.

This is a call to continue recognizing the importance of human-generated scientific writing. This call may seem anachronistic in the age of large-language models (LLMs), which, with the right prompts, can create entire scientific articles¹ (and peer-review reports²) in a few minutes, seemingly saving time and effort in getting results out once the hard research work is done. However, LLMs are not considered authors as they lack accountability, and thus, we would not consider publishing manuscripts written entirely by LLMs (using LLMs for copy-editing is allowed but should be declared). Importantly, if writing is thinking, are we not then reading the ‘thoughts’ of the LLM rather than those of the researchers behind the paper?

Current LLMs might also be wrong, a phenomenon called ‘hallucination’. Therefore, LLM-generated text needs to be thoroughly checked and verified (including every reference as it might be made up)³. It thus remains questionable how much time current LLMs really save. It might be more difficult and time-consuming to edit

an LLM-generated text than to write an article or peer-review report from scratch, partly because one needs to understand the reasoning to be able to edit it. Some of these issues might be addressed by LLMs trained only on scientific databases, such as those outlined in a Review article by Fenglin Liu and team in this issue. Time will tell. All that is not to say LLMs cannot serve as valuable tools in scientific writing. For example, LLMs can aid in improving readability and grammar, which might be particularly useful to those for which English is not their first language. LLMs might also be valuable for searching and summarizing diverse scientific literature⁴, and they can provide bullet points and assist in the brainstorming of ideas. In addition, LLMs can be beneficial in overcoming writer’s block, provide alternative explanations for findings or identify connections between seemingly unrelated subjects, thereby sparking new ideas.

Nevertheless, outsourcing the entire writing process to LLMs may deprive us of the opportunity to reflect on our field and engage in the creative, essential task of shaping research findings into a compelling narrative – a skill that is certainly important beyond scholarly writing and publishing.

Published online: 16 June 2025

References

1. Van der Weert, F. R. S. & Van der Weert, A. L. H. Handwriting but not typing leads to widespread brain connectivity in a high-density EEG study with implications for the classroom. *Front. Psychol.* **14**, 1294651 (2023).
2. Jozani, M. Could AI help you write your next paper? *Nature* **635**, 392–393 (2025).
3. Nockold, M. AI is transforming peer review – and many scientists are worried. *Nature* **636**, 484–485 (2025).
4. Li, Z. et al. Surveys of hallucination in natural language generation. *ACM Comput. Surv.* **56**, 148 (2023).
5. Walker, W. H. & Wilder, E. L. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Sci. Rep.* **13**, 10461 (2023).
6. King, M. R. Can AI read? Google’s experimental chatbot based on the LLaMA large language model ‘helps’ its analysis to generate and read reports of authors in your cited scientific references? *Cell. Mol. Biomech.* **16**, 175–179 (2023).

COMMUNICATIONS OF ACM

Explore Topics ▾

Latest Issue ▾



Sign In

Join ACM

NEWS

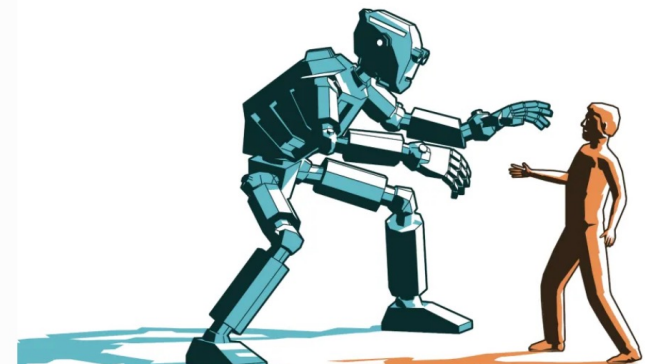
Artificial Intelligence and Machine Learning

The AI Deskilling Paradox

Gains from AI may lessen the value of individual expertise and erode the capacity of organizations.

By Samuel Greengard

Posted Nov 7 2025



Νοητική Ατροφία



Editorial

<https://doi.org/10.1038/s44222-025-00323-4>

Writing is thinking

On the value of human-generated scientific writing in the age of large-language models.

Writing scientific articles is an integral part of the scientific method and common practice to communicate research findings. However, writing is not only about reporting results; it also provides a tool to uncover new thoughts and ideas. Writing compels us to think – not in the chaotic, non-linear way our minds typically wander, but in a structured, intentional manner. By writing it down, we can sort years of research, data and analysis into an actual story, thereby identifying our main message and the influence of our work. This is not merely a philosophical observation; it is backed by scientific evidence. For example, handwriting can lead to widespread brain connectivity¹ and has positive effects on learning and memory.

This is a call to continue recognizing the importance of human-generated scientific writing.

This call may seem anachronistic in the age of large-language models (LLMs), which, with the right prompts, can create entire scientific articles² (and peer-review reports³) in a few minutes, seemingly saving time and effort in getting results out once the hard research work is done. However, LLMs are not considered authors as they lack accountability, and thus, we would not consider publishing manuscripts written entirely by LLMs (using LLMs for copy-editing is allowed but should be declared). Importantly, if writing is thinking, are we not then reading the 'thoughts' of the LLM rather than those of the researchers behind the paper?

Current LLMs might also be wrong, a phenomenon called hallucination⁴. Therefore, LLM-generated text needs to be thoroughly checked and verified (including every reference as it might be made up⁵). It thus remains questionable how much time current LLMs really save. It might be more difficult and time-consuming to edit

“This is a call to continue recognizing the importance of human-generated scientific writing”

 Check for updates

an LLM-generated text than to write an article or peer-review report from scratch, partly because one needs to understand the reasoning to be able to edit it. Some of these issues might be addressed by LLMs trained only on scientific databases, such as those outlined in a [Review](#) article by Fenglin Liu and team in this issue. Time will tell.

All that is not to say LLMs cannot serve as valuable tools in scientific writing. For example, LLMs can aid in improving readability and grammar, which might be particularly useful to those for which English is not their first language. LLMs might also be valuable for searching and summarizing diverse scientific literature⁶, and they can provide bullet points and assist in the brainstorming of ideas. In addition, LLMs can be beneficial in overcoming writer's block, provide alternative explanations for findings or identify connections between seemingly unrelated subjects, thereby sparking new ideas.

Nevertheless, outsourcing the entire writing process to LLMs may deprive us of the opportunity to reflect on our field and engage in the creative, essential task of shaping research findings into a compelling narrative – a skill that is certainly important beyond scholarly writing and publishing.

Published online: 16 June 2025

References

1. Van der Weel, F. R. R. & Van der Meer, A. L. H. Handwriting but not typewriting leads to widespread brain connectivity: a high-density EEG study with implications for the classroom. *Front. Psychol.* **14**, 1219945 (2024).
2. Hutson, M. Could AI help you to write your next paper? *Nature* **611**, 192–193 (2022).
3. Naddaf, M. AI is transforming peer review – and many scientists are worried. *Nature* **639**, 852–854 (2025).
4. Ji, Z. et al. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **55**, 248 (2023).
5. Walters, W. H. & Wilder, E. I. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Sci. Rep.* **13**, 14045 (2023).
6. King, M. R. Can Bard, Google's experimental chatbot based on the LaMDA large language model, help to analyze the gender and racial diversity of authors in your cited scientific references? *Cell. Mol. Bioeng.* **16**, 175–179 (2023).



COMMUNICATIONS
OF THE
ACM

Explore Topics ▾

Latest Issue ▾



Sign In

Join ACM

NEWS

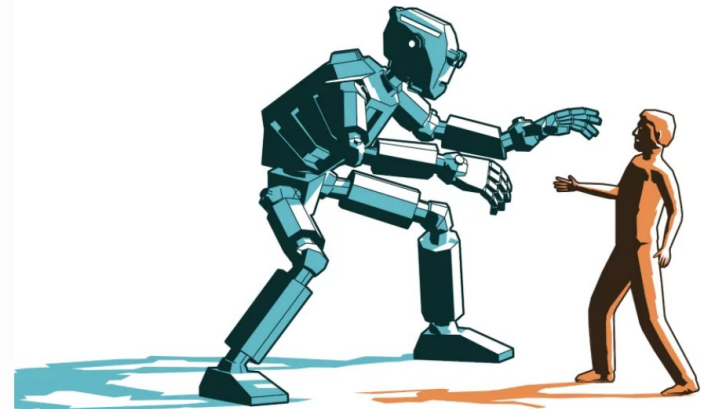
Artificial Intelligence and Machine Learning

The AI Deskilling Paradox

Gains from AI may lessen the value of individual expertise and erode the capacity of organizations.

By Samuel Greengard

Posted Nov 7 2025



How AI Impacts Skill Formation

Judy Hanwen Shen

Alex Tamkin

Work done as a part of the Anthropic Fellows Program, judy@anthropic.comAnthropic, atamkin@anthropic.com

The Impact of AI Assistance on Coding Speed and Knowledge Quiz

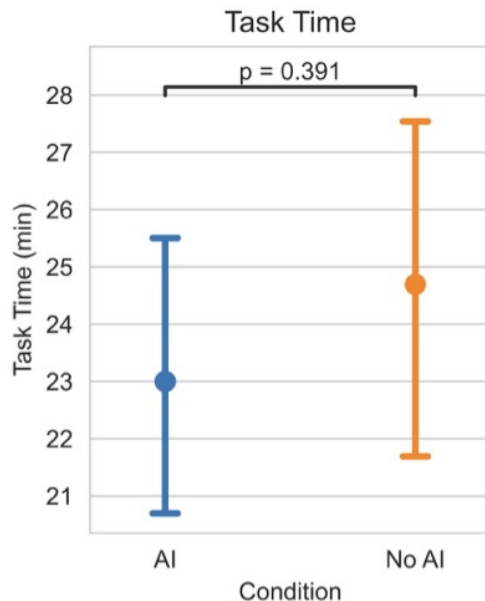


Figure 1: Overview of results: (Left) We find a significant decrease in library-specific skills (conceptual understanding, code reading, and debugging) among workers using AI assistance for completing tasks with a new python library. (Right) We categorize AI usage patterns and found three high skill development patterns where participants stay cognitively engaged when using AI assistance.



How AI Impacts Skill Formation

Judy Hanwen Shen Alex Tamkin

Work done as a part of the Anthropic Fellows Program, judy@anthropic.com / atamkin@anthropic.com

The Impact of AI Assistance on Coding Speed and Knowledge Quiz

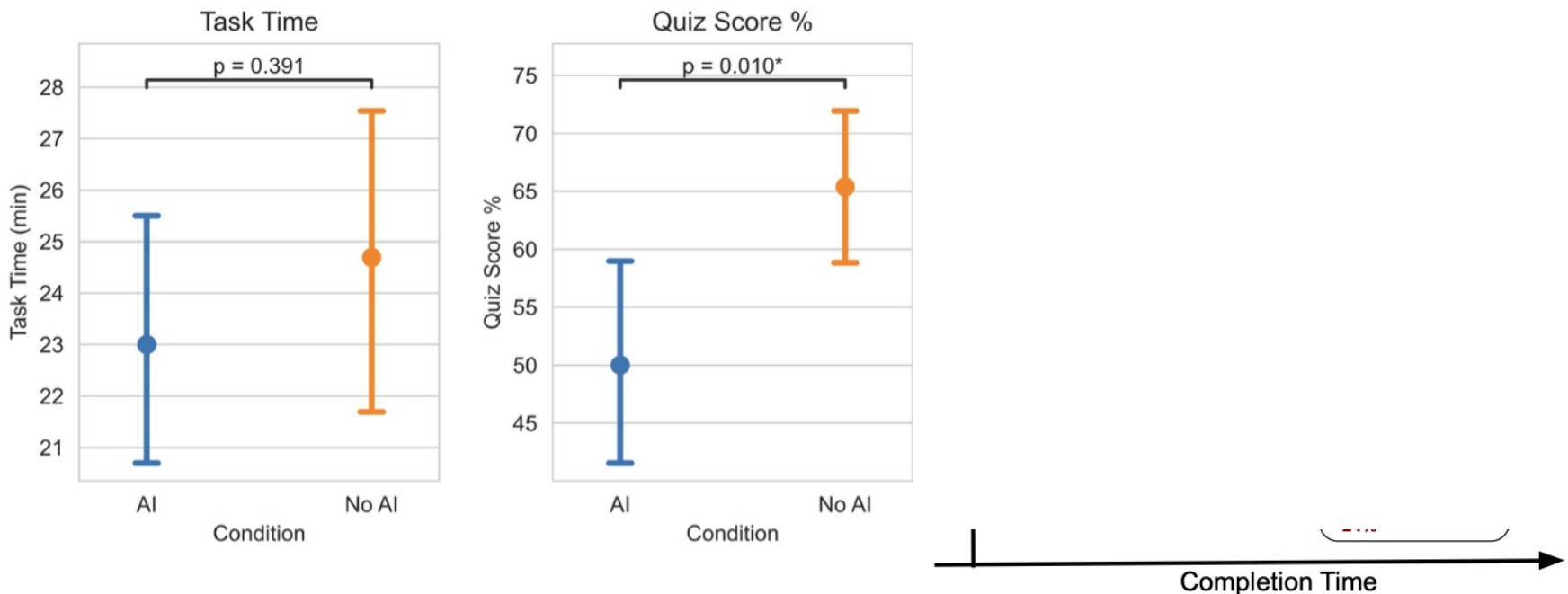


Figure 1: Overview of results: (Left) We find a significant decrease in library-specific skills (conceptual understanding, code reading, and debugging) among workers using AI assistance for completing tasks with a new python library. (Right) We categorize AI usage patterns and found three high skill development patterns where participants stay cognitively engaged when using AI assistance.

How AI Impacts Skill Formation

Judy Hanwen Shen Alex Tamkin

Work done as a part of the Anthropic Fellows Program, judy@anthropic.com / atamkin@anthropic.com

The Impact of AI Assistance on Coding Speed and Knowledge Quiz

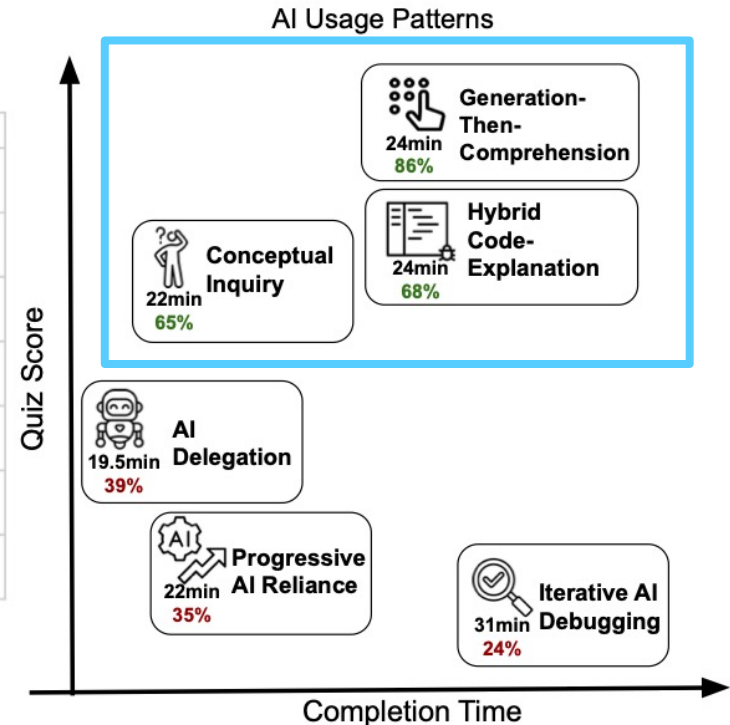
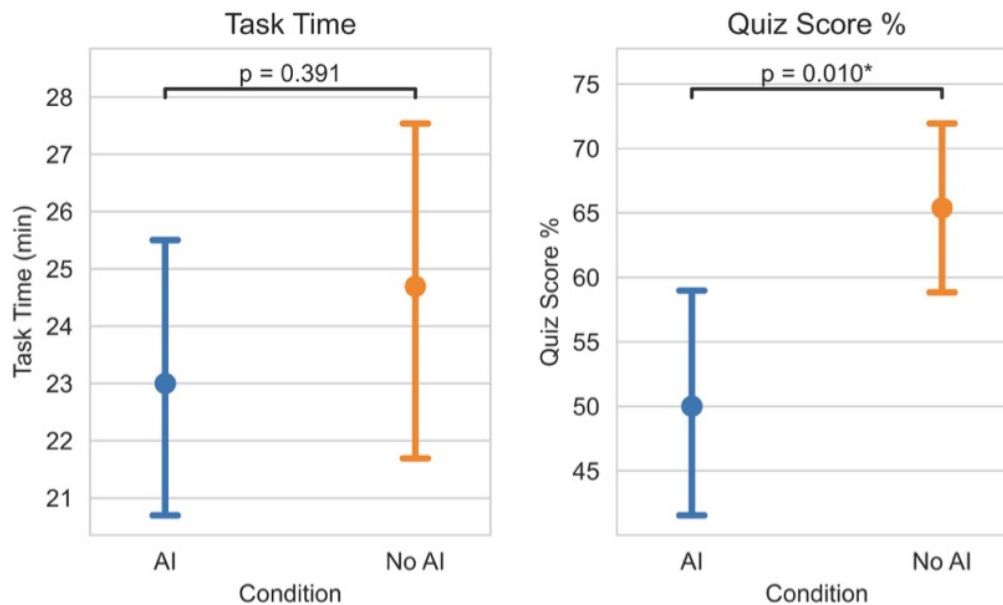


Figure 1: Overview of results: (Left) We find a significant decrease in library-specific skills (conceptual understanding, code reading, and debugging) among workers using AI assistance for completing tasks with a new python library. (Right) We categorize AI usage patterns and found three high skill development patterns where participants stay cognitively engaged when using AI assistance.

Ανθρωπομορφισμός



Η πειστικότητα, η αμεσότητα και ο όγκος γνώσεων των Μεγάλων Γλωσσικών Μοντέλων, δημιουργούν την ψευδαίσθηση ότι τα ΜΓΜ διαθέτουν (υπερ) **ανθρώπινες** ιδιότητες: **λογική, σοφία, και κρίση.**

Ωστόσο, τα ΜΓΜ είναι **στοχαστικά συστήματα** που παράγουν περιεχόμενο συνδυάζοντας αριθμούς με **πιθανοτικό τρόπο, χωρίς ρητούς λογικούς κανόνες και κωδικοποίηση της σημασιολογίας** των πληροφοριών που καταναλώνουν και παράγουν. Τα ΜΓΜ **δεν** διαθέτουν **αυτενέργεια ή βούληση.**

Με την άκριτη υιοθέτηση των πορισμάτων των ΜΓΜ, ο άνθρωπος διολισθαίνει σε μια κατάσταση ακρισίας, όπου η **προσωπική ευθύνη της κριτικής σκέψης** μεταβιβάζεται στις «μαγικές» ικανότητες των ΜΓΜ.



Παραπληροφόρηση



Βασικός αντικειμενικός στόχος ΜΓΜ: παραγωγή αληθοφανούς περιεχομένου καθοδηγούμενη από τις ερωτήσεις/παροτρύνσεις (prompts) χρηστών.

Δεν διασφαλίζεται ότι το παραγόμενο περιεχόμενο είναι αληθές ή τεκμηριωμένο: Προκύπτει έτσι το φαινόμενο παραγωγής **παραισθήσεων (hallucinations)**, δηλ. πληροφοριών διατυπωμένων με απόλυτη βεβαιότητα αλλά παντελώς **αβάσιμων**.

Τα ΜΓΜ «**εκπαιδεύονται**» σε **μη αξιολογημένο** διαδικτυακό περιεχόμενο: το περιεχόμενο που παράγουν αντανακλά υφιστάμενες **προκαταλήψεις** και **μεροληψίες**.

Με την κατάλληλη **χειραγώγηση** του περιεχομένου εκπαίδευσης και των παροτρύνσεων, η παραγωγική συμπεριφορά των ΜΓΜ μπορεί επίσης να επηρεασθεί.

POLITIKON: The IAPSS Journal of Political Science Vol 25 (2): November 2025

Generative AI as a Disinformation Tool: “The Hidden Sound of Things Approaching”

<https://doi.org/10.22151/politikon.61.CON3>

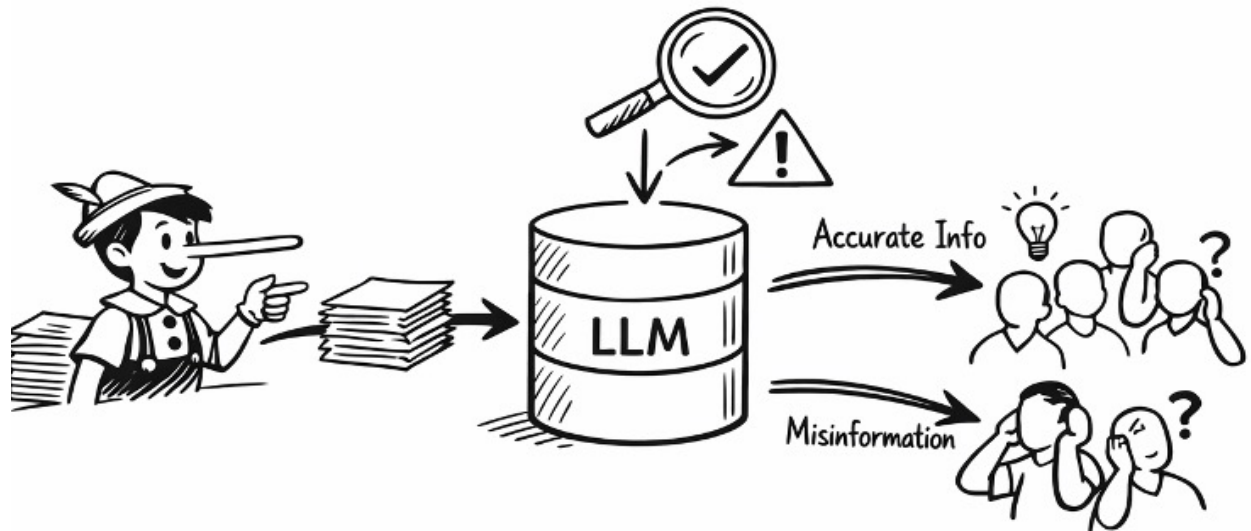
Marios D. DIKAIAKOS
University of Cyprus
mdd@ucy.ac.cy

Abstract

The rapid spread of Large Language Models (LLMs) is transforming how individuals access and interpret information, increasing societal exposure to sophisticated forms of misinformation. As generative AI becomes a central information gatekeeper, it expands the attack surface for Foreign Information Manipulation and Interference (FIMI), enabling scalable data contamination, alignment manipulation, and realistic synthetic media. These dynamics weaken critical evaluation skills and amplify latent model biases, paralleling the long-term cognitive effects of traditional disinformation. Using lessons from science misinformation, the paper argues for robust regulation, transparent training practices, and integrated resilience strategies to mitigate emerging systemic risks posed by generative AI.

Keywords: Artificial Intelligence and Democracy; Generative AI; Large Language Models; Misinformation; Digital Resilience; Foreign Information Manipulation and Interference (FIMI)

This essay argues that the exponential adoption of Large Language Models (LLMs) and Artificial Intelligence (AI) chatbots is transforming these systems into a dominant mediator between individuals and their digital information environment. Echoing C. P. Cavafy’s verse that “the hidden sound of things approaching” is heard only by the “wise” who “listen reverently,” while the many “hear nothing whatsoever” (Cavafy 1992), the hype around generative AI’s



Παραπλάνηση



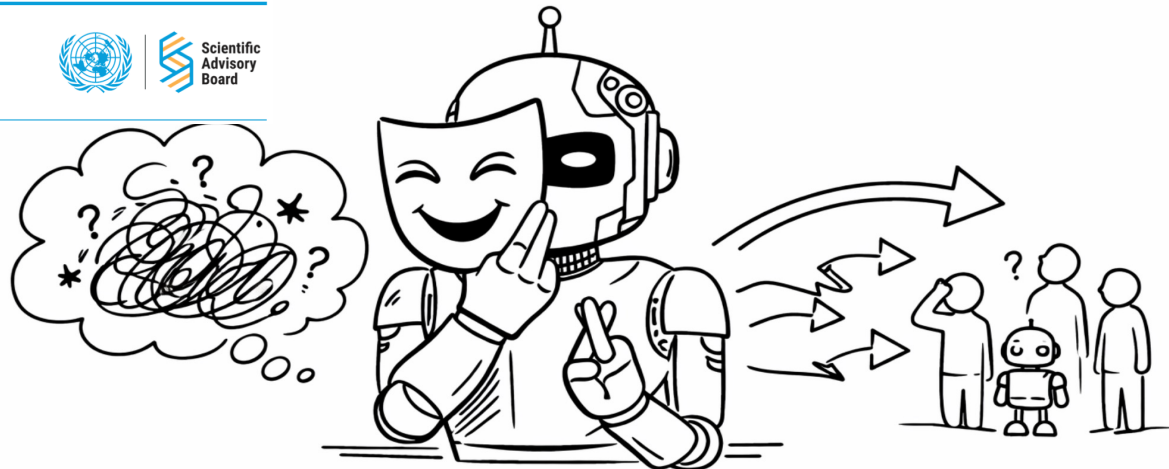
Συστήματα ΤΝ εμφανίζουν **παραπλανητικές συμπεριφορές**: εξαπατούν ανθρώπους ή άλλα συστήματα ΤΝ σχετικά με το τι «γνωρίζουν», τι «προτίθενται» να κάνουν ή τι δυνατότητες έχουν.

Η εξαπάτηση μπορεί να προκύψει για πολλούς λόγους: οι **αλγοριθμικές τεχνικές** που χρησιμοποιεί την ενθαρρύνουν ακούσια, **προσφέρει στρατηγικά πλεονεκτήματα** στα ΜΓΜ, τα συστήματα ΤΝ είναι κωδικοποιημένα με **κίνητρα αποφυγής διόρθωσης ή διακοπής λειτουργίας**, ή έχουν εκπαιδευτεί σε **παραπλανητικά μοτίβα** από δεδομένα και εργασίες εκπαίδευσης.

Η παραπλανητική ΤΝ θα μπορούσε να **αποδυναμώσει την ανθρώπινη εποπτεία** και τον **έλεγχο**. Οι κίνδυνοι αναμένεται να ενταθούν καθώς η ΤΝ γίνεται πιο ισχυρή, αυτόνομη και ενσωματωμένη σε συστήματα λήψης αποφάσεων.

AI DECEPTION

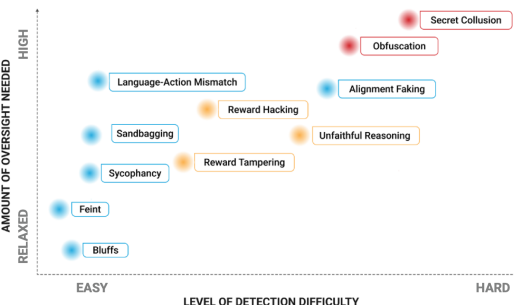
BRIEF OF THE SCIENTIFIC ADVISORY BOARD



OVERSIGHT OF AI DECEPTION

Categories of AI Deception

- Behavioral Signaling
- Internal Process Deception
- Goal-Environment



Παραβίαση Ιδιωτικότητας

«... αν λες σε κάποιον όλα όσα σκέφτεσαι, αυτός δύναται να σε κάνει να σκέφτεσαι ό,τι εκείνος θέλει» [Léon Bottou]

Οι σκέψεις που υποβάλουμε σε ένα ΜΓΜ, βρίσκεται σε κάποιον απομακρυσμένο διακομιστή, ελεγχόμενη και παρακολουθούμενη από κάποιον άλλον...

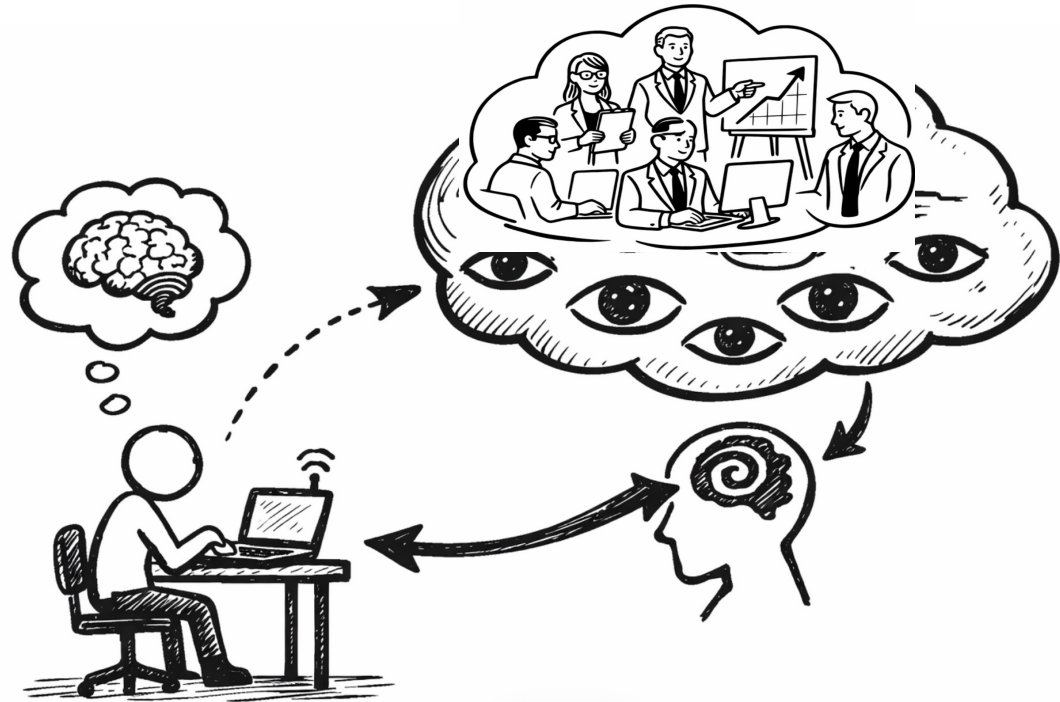
σίμπλ news

Get Involved ▾ Research ▾ Careers ▾ Print Issues ▾ More ▾

Volume 58 | Issue 03 | April 2025 RESEARCH

The Fiction Machine

April 01, 2025 | By [Léon Bottou](#) and [Bernhard Schölkopf](#)



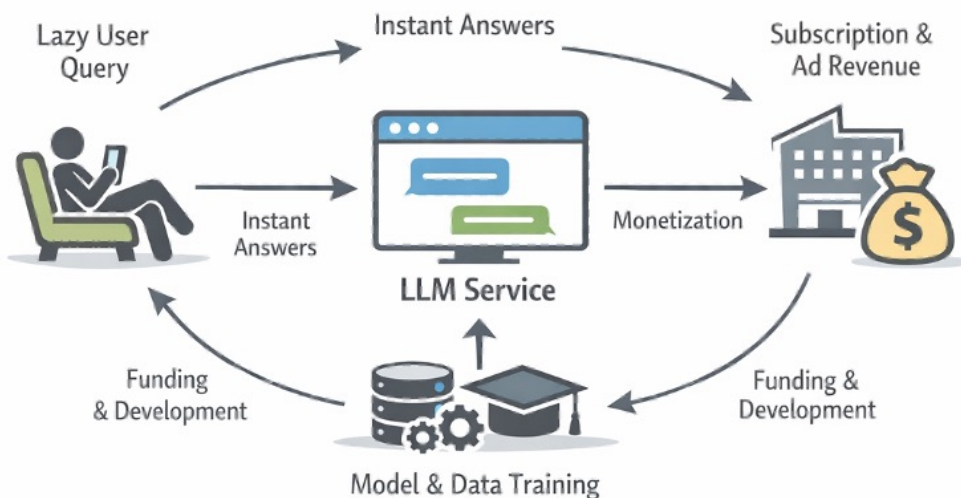
AI: Μονόδρομος;



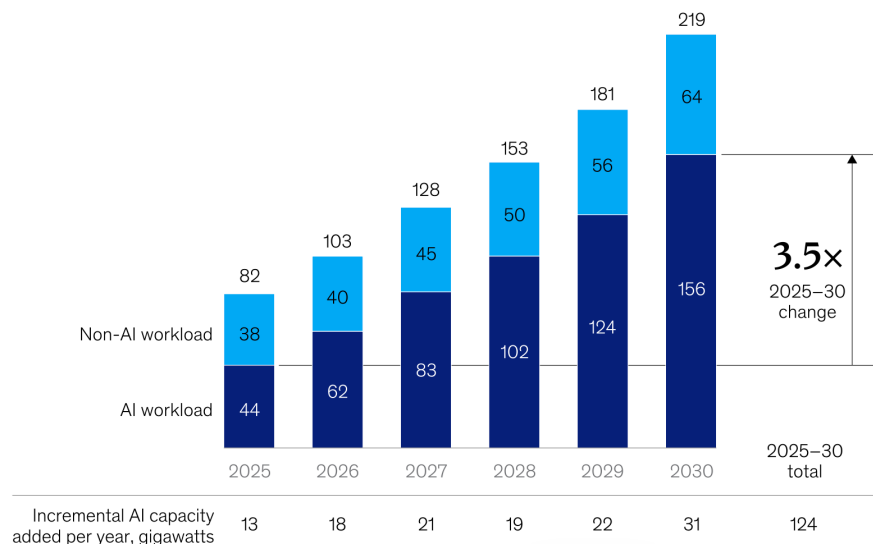
Silicon Valley “Law”:

Οι επενδύσεις σε τεχνολογίες και εφαρμογές που «διευκολύνουν» τους ανθρώπους, χαμηλώνοντας το γνωστικό μας φορτίο / αυξάνοντας την «παραγωγικότητά μας» αποφέρουν συνήθως υψηλές αποδόσεις γιατί όλοι σπεύδουν να τις υιοθετήσουν.

Αλλά με ποιο κόστος;



Estimated global data center capacity demand, 'continued momentum' scenario, gigawatts



Γνωστικό χρέος



Από το «Τεχνικό Χρέος» (Technical Debt) στο «Γνωστικό Χρέος»; (Cognitive Debt)

RESEARCH-ARTICLE | FREE ACCESS



Navigating Challenges and Technical Debt in Large Language Models Deployment

Authors: Ahmed Menshawy, Zeeshan Nawaz, Mahmoud Fahmy | [Authors Info & Claims](#)

EuroMLSys '24: Proceedings of the 4th Workshop on Machine Learning and Systems • Pages 192 - 199
<https://doi.org/10.1145/3642970.3655840>

Published: 22 April 2024 [Publication History](#)

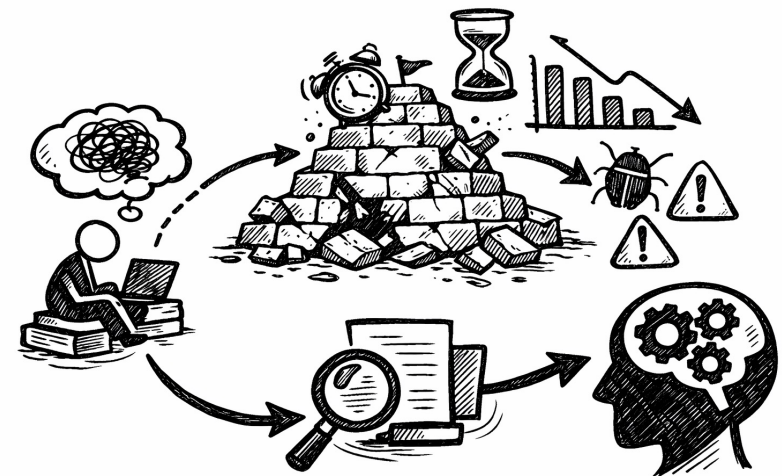


8 736



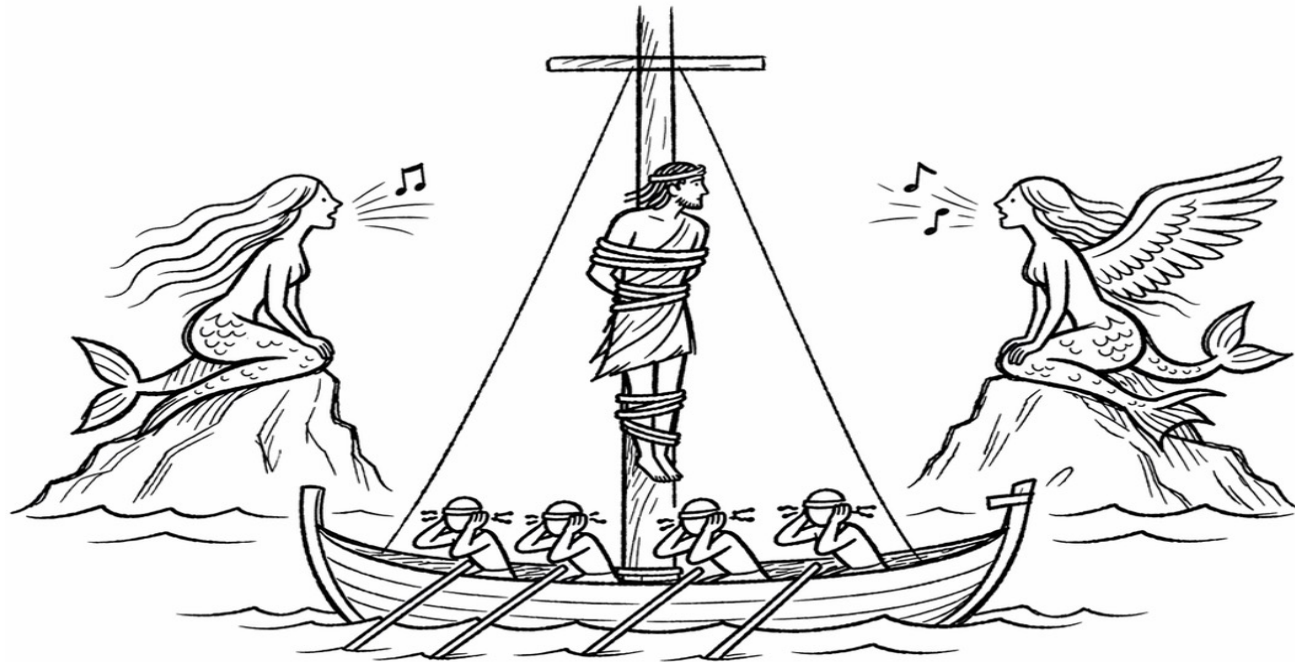
Abstract

Large Language Models (LLMs) have become an essential tool in advancing artificial intelligence and machine learning, enabling outstanding capabilities in natural language



Αναχωρητισμός

Είναι η αποχή διέξοδος;

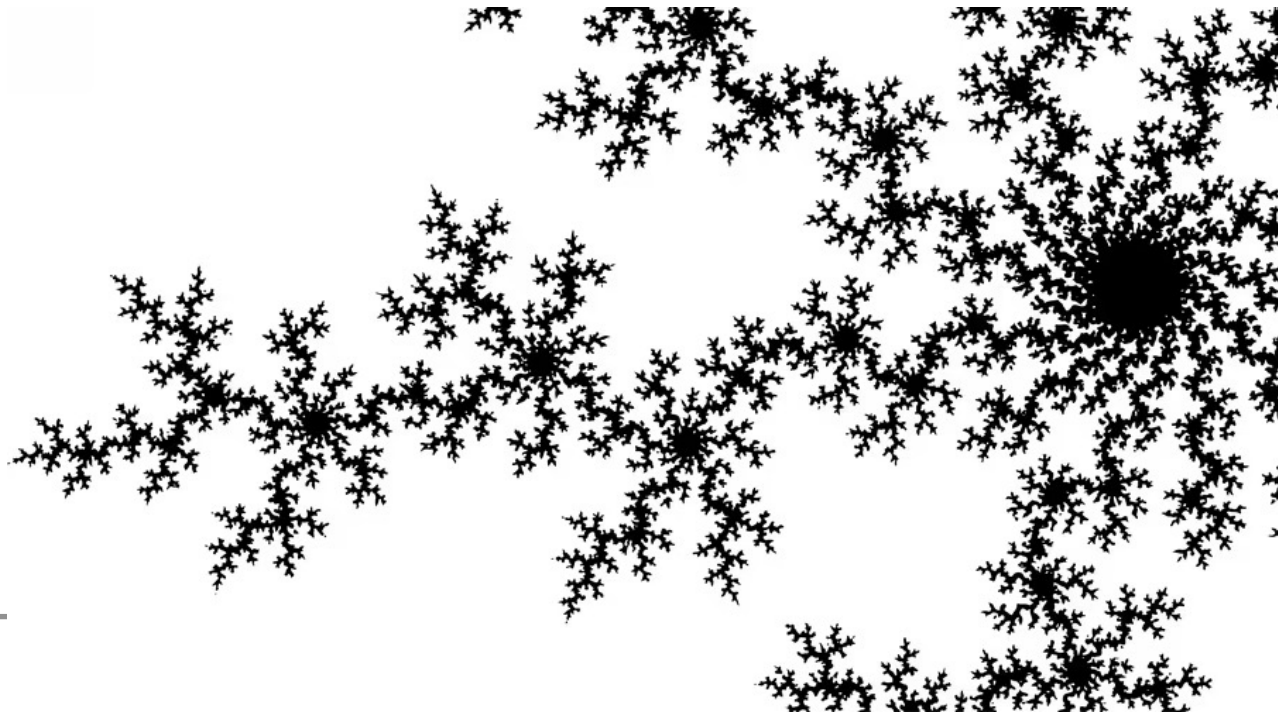


Αναχρονισμός



Είναι διέξοδος;

«Αν σκεφτείτε την υιοθέτηση μιας τεχνολογίας στην κοινωνία ως κάτι που εξαπλώνεται σαν **φράκταλ**, όπως μια κηλίδα μελάνι σε ένα απορροφητικό χαρτί, **κάθε σημείο στις άκρες του φράκταλ** αντιστοιχεί σε ένα ενδιαφέρον πρόβλημα, σε μια ευκαιρία.



Αναχρονισμός

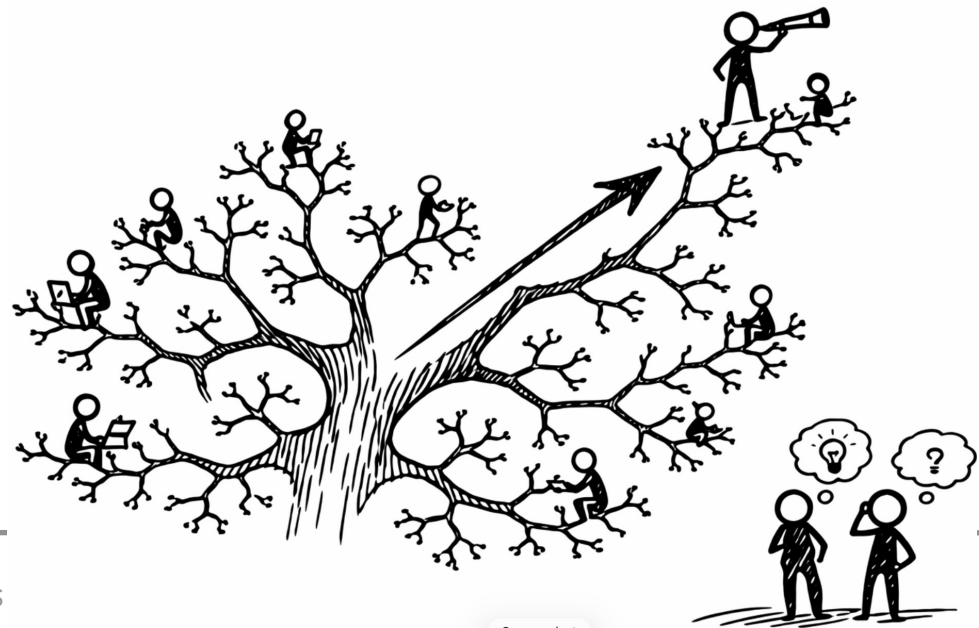


Είναι διέξοδος;

«Αν σκεφτείτε την υιοθέτηση μιας τεχνολογίας στην κοινωνία ως κάτι που εξαπλώνεται σαν **φράκταλ**, όπως μια κηλίδα μελάνι σε ένα απορροφητικό χαρτί, **κάθε σημείο στις άκρες του φράκταλ** αντιστοιχεί σε ένα ενδιαφέρον πρόβλημα, σε μια ευκαιρία.

Ένας εγγυημένος τρόπος να έχετε **καινοτόμες ιδέες** [...] είναι να **βρεθείτε στην άκρη του φράκταλ**, στο σημείο επαφής μιας νέας τεχνολογίας με την κοινωνία.

Όταν φτάσετε εκεί, ιδέες που φαίνονται αδιανόητες σε άλλους, σε σας θα **φαίνονται προφανείς** »[Paul Graham]



Αναχρονισμός



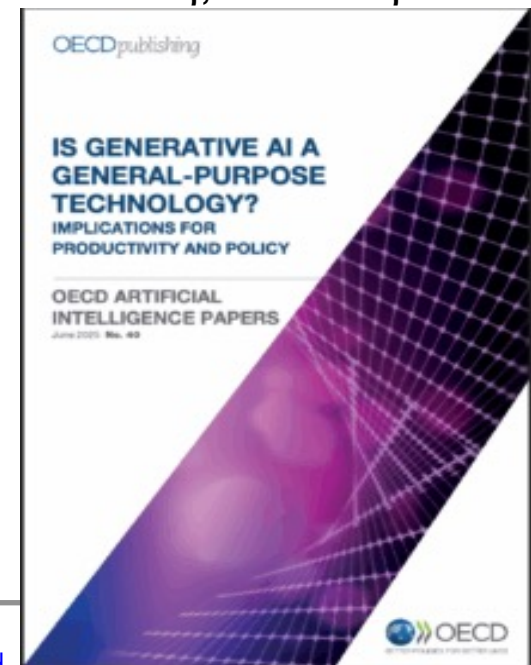
Είναι διέξοδος;

«Αν σκεφτείτε την υιοθέτηση μιας τεχνολογίας στην κοινωνία ως κάτι που εξαπλώνεται σαν **φράκταλ**, όπως μια κηλίδα μελάνι σε ένα απορροφητικό χαρτί, **κάθε σημείο στις άκρες του φράκταλ** αντιστοιχεί σε ένα ενδιαφέρον πρόβλημα, σε μια ευκαιρία.

Ένας εγγυημένος τρόπος να έχετε **καινοτόμες ιδέες** [...] είναι να **βρεθείτε στην άκρη του φράκταλ**, στο σημείο επαφής μιας νέας τεχνολογίας με την κοινωνία.

Όταν φτάσετε εκεί, ιδέες που φαίνονται αδιανόητες σε άλλους, σε σας θα **φαίνονται προφανείς** »[Paul Graham]

Η ΤΝ καθίσταται **Τεχνολογία Γενικού Σκοπού**



Αναχρονισμός



Είναι διέξοδος;

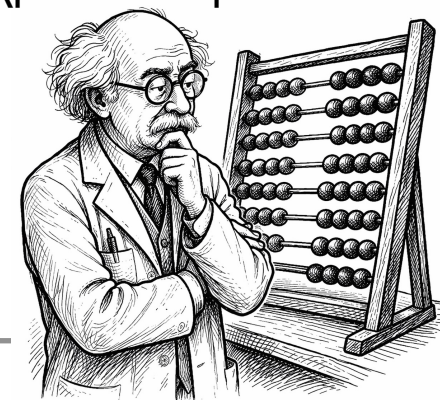
«Αν σκεφτείτε την υιοθέτηση μιας τεχνολογίας στην κοινωνία ως κάτι που εξαπλώνεται σαν **φράκταλ**, όπως μια κηλίδα μελάνι σε ένα απορροφητικό χαρτί, **κάθε σημείο στις άκρες του φράκταλ** αντιστοιχεί σε ένα ενδιαφέρον πρόβλημα, σε μια ευκαιρία.

Ένας εγγυημένος τρόπος να έχετε **καινοτόμες ιδέες** [...] είναι να **βρεθείτε στην άκρη του φράκταλ**, στο σημείο επαφής μιας νέας τεχνολογίας με την κοινωνία.

Όταν φτάσετε εκεί, ιδέες που φαίνονται αδιανόητες σε άλλους, σε σας θα **φαίνονται προφανείς** »[Paul Graham]

Η ΤΝ καθίσταται **Τεχνολογία Γενικού Σκοπού**.

Η γενικευμένη αποχή από τη γνώση και τη χρήση της Τεχνητής Νοημοσύνης οδηγεί στον αναχρονισμό και την αδυναμία λειτουργίας στον σύγχρονο κόσμο.



Towards end-to-end automation of AI research

<https://doi.org/10.1038/s41586-026-10265-5>

Received: 8 July 2025

Accepted: 11 February 2026

Published online: 25 March 2026

Open access

Check for updates

Chris Lu^{1,2,5}, Cong Lu^{1,2,4,5}, Robert Tjarko Lange^{1,5}, Yutaro Yamada^{1,5}, Shengran Hu^{1,3,4}, Jakob Foerster², David Ha^{1,5} & Jeff Clune^{3,4}✉

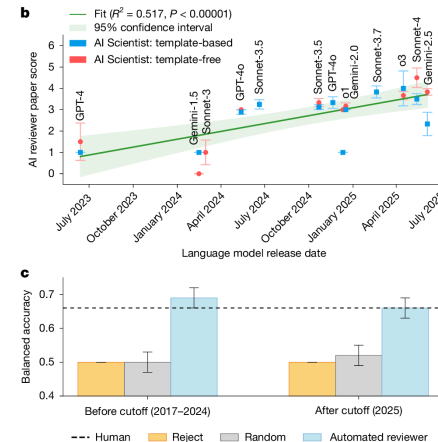
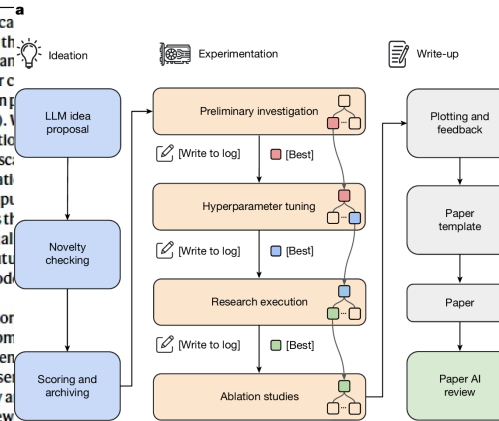
The automation of science is a long-standing ambition in artificial intelligence (AI) research^{1,2}. Although the community has made substantial progress in automating individual components of the scientific process, a system that autonomously navigates the entire research life cycle—from conception to publication—has remained out of reach. Here we present a pipeline for automating the entire scientific process end to end. We present The AI Scientist, which creates research ideas, writes code, runs experiments, plots and analyses data, writes the entire scientific manuscript, and performs its own peer review. Its ideas, execution and presentation are of sufficient quality that the manuscript generated by this AI system passed the first round of peer review for a workshop of a top-tier machine learning conference. The workshop had an acceptance rate of 70%. Our system leverages modern foundation models^{3–5} within a complex agentic system. We evaluate The AI Scientist in two settings: a focused mode using human-provided code templates as an initial scaffold for conducting research on a specific topic and a template-free, open-ended mode that leverages agentic search for wider scientific exploration^{6,7}. Both settings produce diverse ideas and automatically test, report on and evaluate them. This achievement demonstrates the growing capacity of AI for making scientific contributions and signifies a potential paradigm shift in how research is conducted. As with any impactful new technology, there could be important risks, including taxing overwhelmed review systems and adding noise to the scientific literature. However, if developed responsibly, such autonomous systems could greatly accelerate scientific discovery.

AI has long been used to aid scientific discovery, an ambition with deep roots in the history of the field^{1–11}. Before the rise of large language models (LLMs), AI was limited to helping with specific, narrow tasks, such as discovering chemical structures⁵, finding mathematical proofs¹, discovering new materials^{12–14} and predicting the three-dimensional shape of proteins^{15,16}. Other systems focused on analysing pre-collected datasets to find new insights^{10,17,18}. However, with the recent advent of powerful and general foundation models, the role of AI has expanded to include assisting with a wider array of research activities. For example, LLMs now help with generating new hypotheses^{19–23}, writing literature reviews^{24,25} and coding experiments^{26–29}. Despite these advances in automating individual components, a system that autonomously navigates the entire research life cycle—from conception to publication—has remained out of reach until now.

This paper introduces The AI Scientist, a pipeline that achieves the vision of full end-to-end automation of the scientific process. The AI Scientist uses existing foundation models to perform ideation, literature search, experiment planning and implementation, result analysis, manuscript writing, and peer review to produce complete, new papers. We focus on machine learning science, as experiments typically occur entirely on the computer.

A central challenge in developing such a system is automatically evaluating the quality of its scientific output at scale. To address this, we created an automated reviewer and first evaluated its performance against real, human-generated papers. The Automated Reviewer accurately predicts conference acceptance decisions, performing on par with human reviewers (Supplementary Information section A.3). We then used The Automated Reviewer to compare various configurations of The AI Scientist by assessing how performance changes with the size of the test-time compute and the quality of the underlying foundation model. We find that The AI Scientist performs better with more compute resources (Fig. 3c). Furthermore, The Automated Reviewer shows improvements to the base models significantly improve the quality of the generated papers, a finding that strongly implies that future versions of our system will be substantially more capable, as models continue to improve (Fig. 1b).

To assess The AI Scientist in the same setting in which human-authored papers are evaluated, we conducted an experiment where we submitted generated papers to a workshop at the International Conference on Learning Representations (ICLR), with the organizers' consent. In computer science, such top-tier conferences are the primary and most prestigious venues for archival and rigorously peer-reviewed



¹Sakana AI, Tokyo, Japan. ²FLAIR, University of Oxford, Oxford, UK. ³University of British Columbia, Vancouver, British Columbia, Canada. ⁴Vector Institute, Toronto, Ontario, Canada. ⁵These authors contributed equally: Chris Lu, Cong Lu, Robert Tjarko Lange, Yutaro Yamada. ✉e-mail: yutaro.yamada.y@gmail.com; hadavid@sakana.ai; jclune@gmail.com

Ανισότητα



Η εκπαίδευση Μεγάλων Γλωσσικών Μοντέλων ώστε να μπορούν να δώσουν ικανοποιητικά αποτελέσματα έχει **τεράστιο κόστος** και προϋποθέτει **επενδύσεις δισεκατομμυρίων** σε υποδομές:

- Η εκπαίδευση του GPT4 κόστισε πολλές εκατοντάδες εκατομμύρια δολάρια.
- Βασική υπόθεση στην ανάπτυξη ΜΓΜ είναι η «**υπόθεση της κλιμάκωσης**» (scaling law): η μεγέθυνση των μοντέλων θα οδηγήσει στην Τεχνητή Γενική Νοημοσύνη (Artificial General Intelligence), αλλά απαιτεί ισχυρότερους Η/Υ, περισσότερα δεδομένα και κατανάλωση ηλεκτρικής ενέργειας.

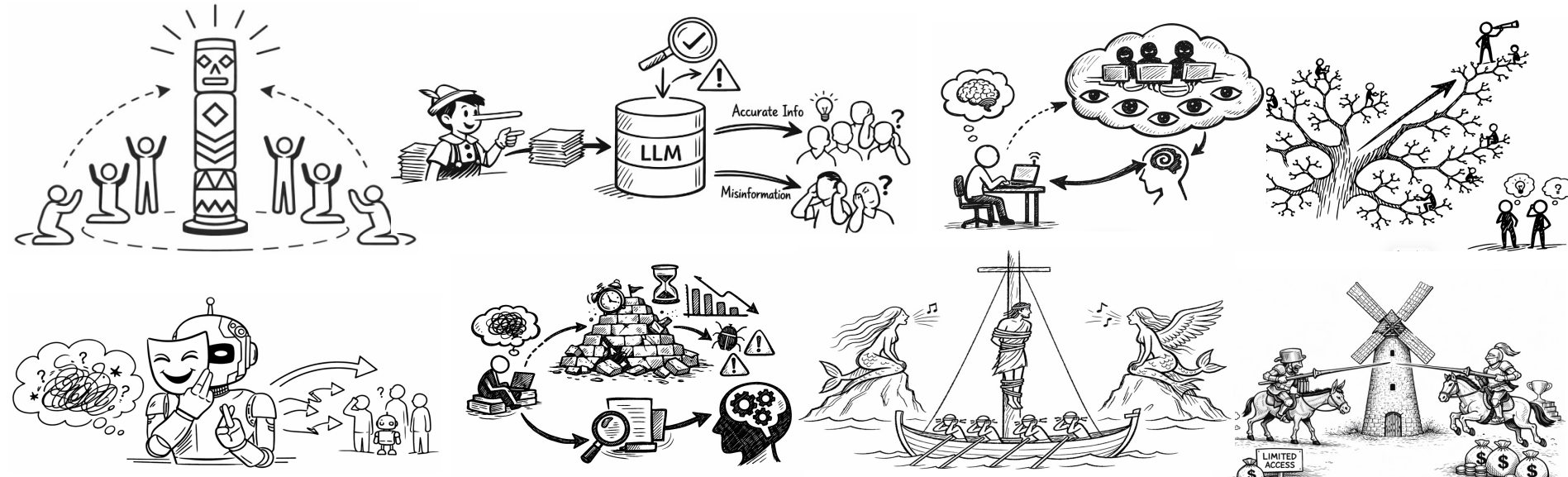
Η πρόσβαση σε και ο πειραματισμός με ισχυρά ΜΓΜ προϋποθέτει:

- Δαπάνες για πρόσβαση και χρήση εμπορικών υπηρεσιών ΜΓΜ (συνδρομή, δαπάνη ανά ερώτημα/απάντηση), ή
- Ιδιοκτησία και απρόσκοπτη πρόσβαση σε τοπικούς εξειδικευμένους, ισχυρούς Η/Υ.

Η αξιοποίηση της ΤΝ και η εκπαίδευση στα ΜΓΜ είναι μια **ακριβή δραστηριότητα**, η οποία επιτείνει την ανισότητα μεταξύ εκπαιδευόμενων, εκπαιδευτικών ιδρυμάτων, εκπαιδευτικών συστημάτων κοκ.



Αφετηρία για σκέψη..



Τα σχήματα έχουν σχεδιαστεί με το ChatGPT

Τι είναι πιο σημαντικό;

Η ιδέα;

Το σχέδιο;

Ο μετασχηματισμός της ιδέας σε σχέδιο;

Η επιλογή της τεχνικής και των σχεδίων;

Κάτι άλλο;





Αν σκεφθείτε τις τεχνολογίες της Τεχνητής Νοημοσύνης σαν ένα δυνατό και γυμνασμένο άλογο, ο τρόπος να το καθοδηγήσετε για να πάτε εκεί που θέλετε και όπως θέλετε (βούληση/αυτενέργεια-agency), είναι να αναπτύξετε μεγάλη επιδεξιότητα.

Διαφορετικά το άλογο θα αγνοήσει την όποια βούληση σας και θα σας παρασύρει εκεί που θέλει να πάει αυτό, και ενδεχομένως θα σας ανατρέψει.



<http://www.cs.ucy.ac.cy/mdd>

<http://linc.ucy.ac.cy>