

Między wsparciem a zagrożeniem: ciemna strona AI

JOANNA MYTNIK, CENTRUM NOWOCZESNEJ EDUKACJI POLITECHNIKI GDAŃSKIEJ

1. Jak uczniowie i studenci używają GenAI i jakie emocje im towarzyszą?
2. **Zagrożenia poznawcze** (delegowanie myślenia)
3. **Pochlebność i ugodowość LLM**
4. **AI companion** (towarzysze AI)
5. **Zagrożenia informacyjne**
 - Fałszywe treści**
 - mis-information
 - dezinformacja
 - mal-information
 - Techniki i narzędzia (inżynieria społeczna)**
 - deepfake
 - manipulacja (clickbajt, moral outrage)
 - phishing i inne techniki inżynierii społecznej
6. **Odporność psychiczna** (edukacja, modelowanie, sposoby budowania)

Zmiana paradygmatu

Paradygmat to rama myślenia, zestaw założeń, które uznajemy za oczywiste, nawet ich nie zauważając.

Zmiana paradygmatu

Przez wieki obowiązywał paradygmat, w którym:

- informacja była dobrem trudnym do zdobycia (rzadkim)
- uczelnia była strażnikiem prawdy i autorytetu
- uczenie się polegało na przyswajaniu tego, co już uznane za wiarygodne

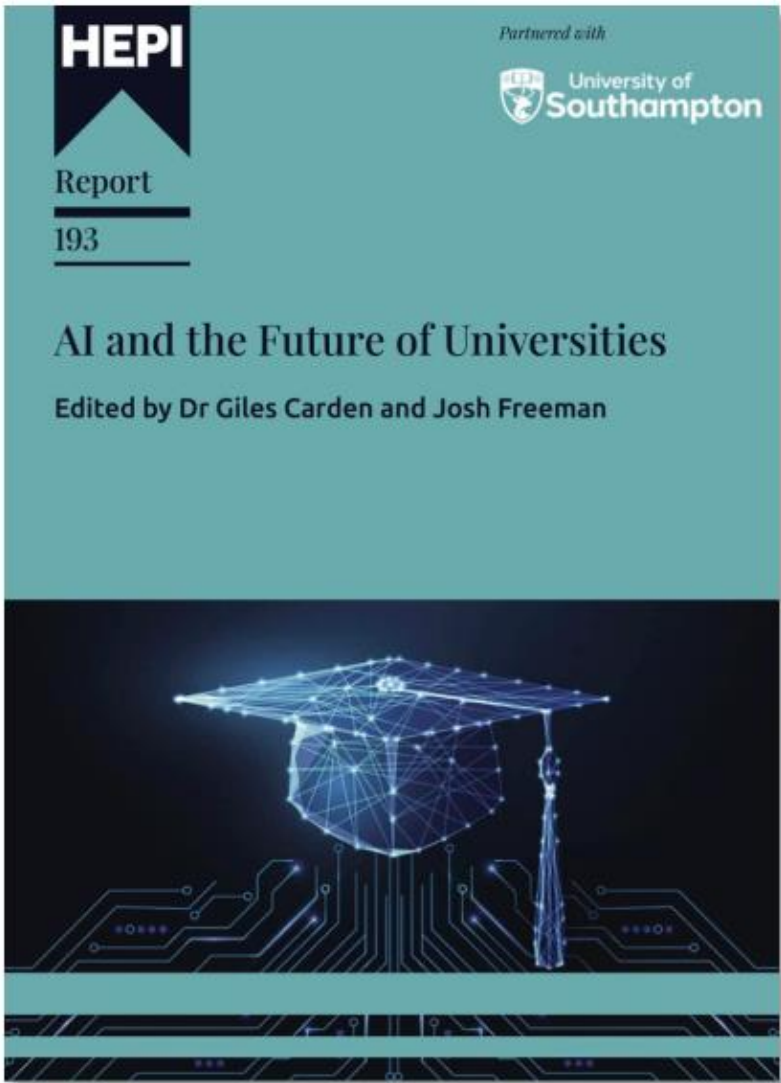
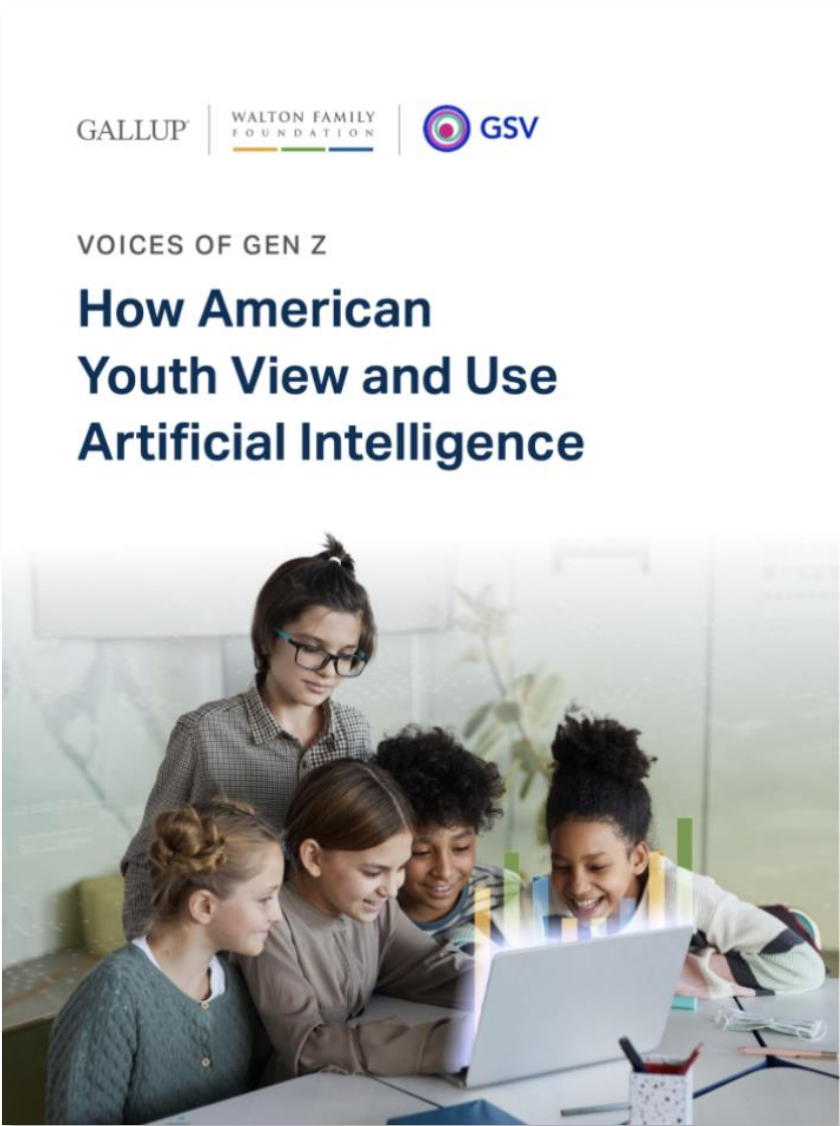
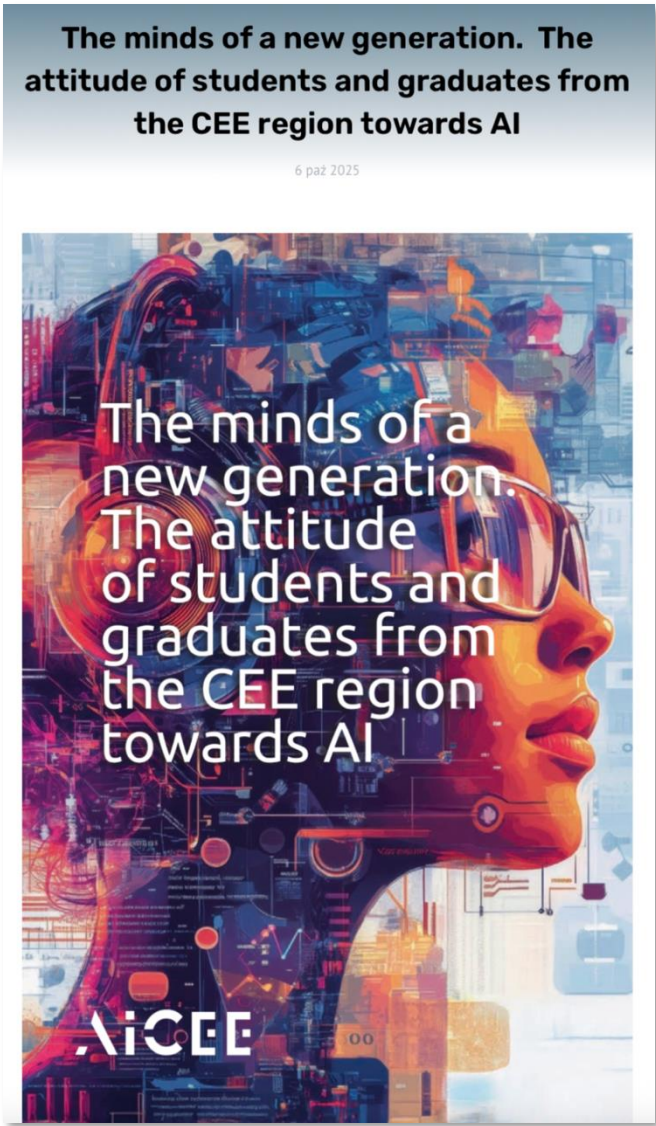
Al wszystko to narusza, bo:

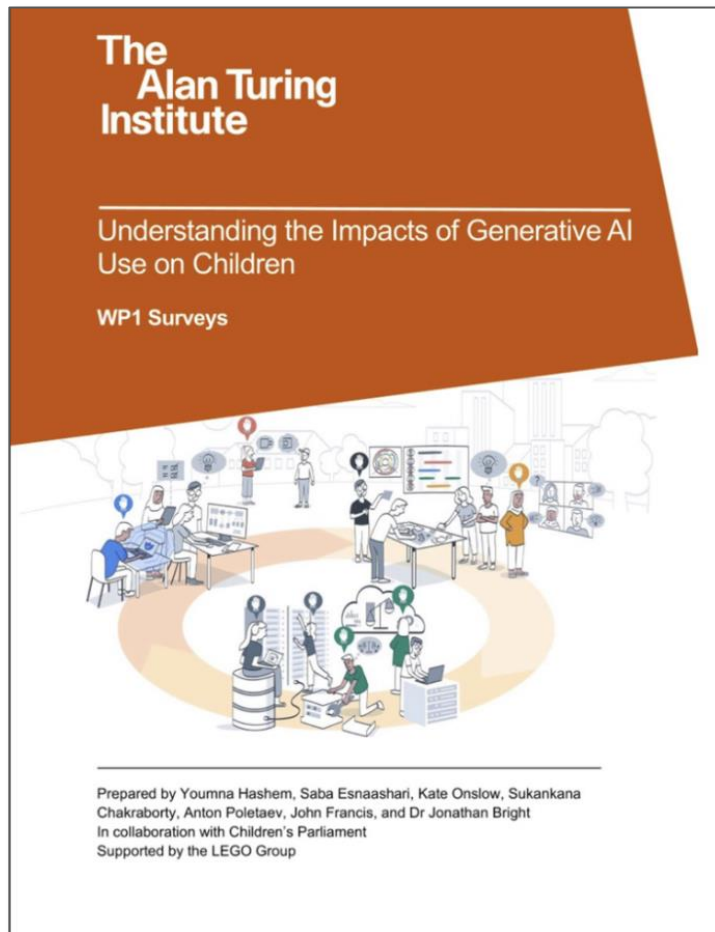
- **informacja stała się łatwo dostępna**
- **definicja autorstwa się rozmywa**
- **granica zaufania przesuwana się**

Zmiana paradygmatu

To oznacza, że zmienia się nie tylko technologia, ale sam **sposób rozumienia, czym jest wiedza**, jak powstaje i kto ma prawo ją interpretować.

Uniwersytet w epoce post-AI

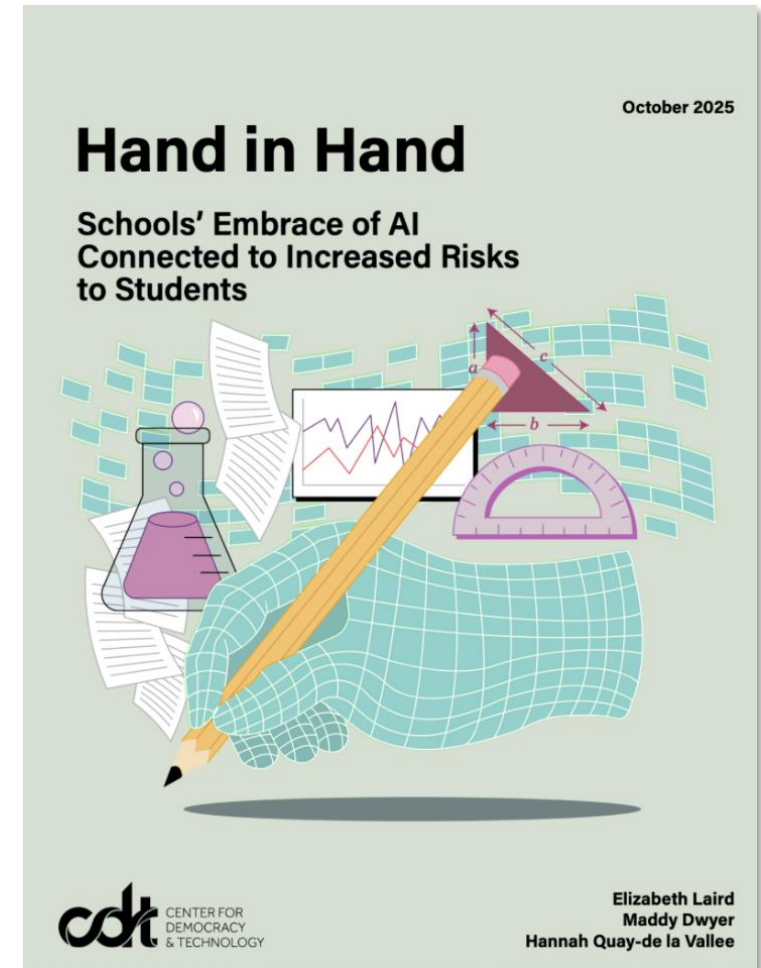




Badanie: 2600 dzieci (8-17 lat), 1200 rodziców, 800 nauczycieli z Wielkiej Brytanii



Badani: uczniowie 7 i 8 klasy szkoły podstawowej oraz 1 i 2 klasy szkoły ponadpodstawowej.



Uczniowie szkół średnich (klasy 9-12): 1 030
nauczyciele klas średnich i gimnazjalnych:
806 rodzice uczniów: 1018 (USA, 2025)

Uczniowie apelują o wskazówki:

są zaniepokojeni tym, że **AI zachęca do kopiowania** i generuje **stronnicze treści**.

Wyraźnie domagają się **jaśniejszych wytycznych instytucjonalnych** i większego **wsparcia dla nauczycieli** w zakresie odpowiedzialnego korzystania ze AI.

AI & Education

OUP Report: Teaching the AI-Native Generation

Over half of UK pupils unable to spot AI misinformation as teenagers call for more guidance from schools

- UK teenagers want support from teachers to identify trustworthy AI content
- 60 percent of pupils worry AI tools encourage copying rather than original work
- Young people believe AI helps skills development but outline concerns
- Oxford University Press unveils AI and Education Site to support schools

Wednesday 15th October

Fewer than half of UK pupils (47 percent) feel confident identifying accurate AI-generated information, according to new research by Oxford University Press (OUP), surveying 2,000 students aged 13–18 across the UK. One third (32 percent) stated they can't tell if AI content is true, whilst a fifth (21 percent) were unsure. The findings also showed that 8 in 10 young people use AI tools in their schoolwork.

Students want more clarity as a third feel their teachers lack confidence when using AI tools

The research revealed that almost half (48 percent) of pupils want support from their teachers to help them understand what content generated by AI is trustworthy and reliable, with 51 percent of pupils calling for more clarity on when they should be using AI tools in their schoolwork. A further third of students believe their teachers are not confident in using AI tools in lessons and one in three pupils

2000 uczniów z Wielkiej Brytanii, wiek 13-18 lat

Uczelnie wysyłają niejasne i niespójne sygnały.

Studenci czują, że uczelnie narzucają surowe zasady, jednocześnie nie znają ich, polegają na plotkach i domysłach („detektory” AI).

W efekcie **boją się używać** AI jednocześnie czując, że **muszą jej używać**.

Próbują balansować między efektywnością a ryzykiem.

Studenci uważają, że ich praca powinna pochodzić od nich i to jest ważniejsze niż reguły formalne.

„**Własność pracy**” oznacza dla studentów:

- że pomysły są ich,
- że proces jest ich,
- że rozumieją to, co oddają,
- że nie czują się „zastąpieni”.

HIGHER EDUCATION RESEARCH & DEVELOPMENT
<https://doi.org/10.1080/07294360.2025.2580616>

 **Routledge**
Taylor & Francis Group

 OPEN ACCESS  Check for updates

Time, emotions and moral judgements: how university students position GenAI within their study

Margaret Bearman ^a, T. Fawns^b, T. Corbin^a, M. Henderson^c, Y. Liang^d, G. Oberg^d, J. Walton^a and K. E. Matthews^d

^aCentre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, Australia;

^bMonash Education Academy, Monash University, Melbourne, Australia; ^cFaculty of Education, Monash University, Melbourne, Australia; ^dOffice of the Deputy Vice-Chancellor (Academic), University of Queensland, Brisbane, Australia

ABSTRACT

The emergence of Generative AI (GenAI) in higher education has prompted considerable discussion within the research community. Despite their centrality, students' perspectives remain underexplored. We investigated how students position GenAI in relation to their studies, conducting focus groups with 79 students across four Australian universities. Taking a sociotechnical stance and employing reflexive thematic analysis, we identified three primary themes: (1) studying with and without GenAI; (2) mixing messages and assumptions; and (3) 'coming from me': self-trust and resistance to dependency. Crossing these themes were axial threads of time, emotions, and moral judgement. Our findings illuminate a complex, dynamic and uncertain landscape of relationships in which students prioritise their developing values and moral positions over institutional messaging. Amongst diverse and changing practices, students generally conveyed a sense that their educational work should come from, and be owned by, themselves. This concept of ownership could productively inform academic integrity discussions. Institutions may need to move beyond rules and restrictions, to foster environments where students can develop nuanced relationships with GenAI whilst maintaining academic autonomy. This research demonstrates the value of centring student voices in shaping scholarly practice and policy, particularly in contexts of significant technological change.

ARTICLE HISTORY

Received 4 November 2024
Accepted 12 October 2025

KEYWORDS

Artificial intelligence; generative artificial intelligence; sociotechnical; qualitative methods; academic integrity; higher education

Studenci obawiają się uzależnienia od AI.

Wielu mówi o:

- **lęku** przed utratą umiejętności,
- **obawie**, że AI „rozleniwi”,
- **strachu**, że w egzaminie zostaną bez narzędzi i „to się wyda”.

Studenci chcą przewodnictwa, ale nie kontroli.

Badanie pokazuje, że **studenci potrzebują**:

- jasnych zasad,
- wskazówek, jak uczyć się z AI,
- zrozumieć, czego uczelnia od nich oczekuje.

Uczelnie są postrzegane jako **spóźnione, niepewne, często milczące.**

HIGHER EDUCATION RESEARCH & DEVELOPMENT
<https://doi.org/10.1080/07294360.2025.2580616>

 **Routledge**
Taylor & Francis Group

 OPEN ACCESS  Check for updates

Time, emotions and moral judgements: how university students position GenAI within their study

Margaret Bearman ^a, T. Fawns^b, T. Corbin^a, M. Henderson^c, Y. Liang^d, G. Oberg^d, J. Walton^a and K. E. Matthews^d

^aCentre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, Australia;

^bMonash Education Academy, Monash University, Melbourne, Australia; ^cFaculty of Education, Monash University, Melbourne, Australia; ^dOffice of the Deputy Vice-Chancellor (Academic), University of Queensland, Brisbane, Australia

ABSTRACT

The emergence of Generative AI (GenAI) in higher education has prompted considerable discussion within the research community. Despite their centrality, students' perspectives remain underexplored. We investigated how students position GenAI in relation to their studies, conducting focus groups with 79 students across four Australian universities. Taking a sociotechnical stance and employing reflexive thematic analysis, we identified three primary themes: (1) studying with and without GenAI; (2) mixing messages and assumptions; and (3) 'coming from me': self-trust and resistance to dependency. Crossing these themes were axial threads of time, emotions, and moral judgement. Our findings illuminate a complex, dynamic and uncertain landscape of relationships in which students prioritise their developing values and moral positions over institutional messaging. Amongst diverse and changing practices, students generally conveyed a sense that their educational work should come from, and be owned by, themselves. This concept of ownership could productively inform academic integrity discussions. Institutions may need to move beyond rules and restrictions, to foster environments where students can develop nuanced relationships with GenAI whilst maintaining academic autonomy. This research demonstrates the value of centring student voices in shaping scholarly practice and policy, particularly in contexts of significant technological change.

ARTICLE HISTORY

Received 4 November 2024
Accepted 12 October 2025

KEYWORDS

Artificial intelligence; generative artificial intelligence; sociotechnical; qualitative methods; academic integrity; higher education

Uczelnie są postrzegane jako spóźnione, niepewne,
często milczące.

Zagrożenia poznawcze

- dostęp do chatGPT znacznie **podnosi wydajność** osób uczących się (48% w przypadku chatGPT wersja podstawowa i 127% dla GPT-tutor)
- po odebraniu tego dostępu: osoby korzystające z **chatGPT osiągają gorsze wyniki niż osoby nigdy nie korzystające ze wsparcia AI (17%)**
- skutki te są **mniej negatywne** w przypadku pracy z GPT-tutor (dzięki obecnym w nim zabezpieczeniom)
- wyniki wskazują na możliwy **negatywny** wpływ nadmiernego polegania na AI na trwałość uczenia się, szczególnie u osób o niskim poziomie samoregulacji i krytycznego myślenia.

Generative AI Can Harm Learning

[The Wharton School Research Paper](#)

59 Pages • Posted: 18 Jul 2024

[Hamsa Bastani](#)

University of Pennsylvania - The Wharton School

[Osbert Bastani](#)

University of Pennsylvania - Department of Computer and Information Science

[Alp Sungu](#)

University of Pennsylvania

[Haosen Ge](#)

University of Pennsylvania - The Wharton School

[Özge Kabakcı](#)

Budapest British International School

[Rei Mariman](#)

Independent; Independent

Date Written: July 15, 2024

Abstract

Generative artificial intelligence (AI) is poised to revolutionize how humans work, and has already demonstrated promise in significantly improving human productivity. However, a key remaining question is how generative AI affects *learning*, namely, how humans acquire new skills as they perform tasks. This kind of skill learning is critical to long-term productivity gains, especially in domains where generative AI is fallible and human experts must check its outputs. We study the impact of generative AI, specifically OpenAI's GPT-4, on human learning in the context of math classes at a high school. In a field experiment involving nearly a thousand students, we have deployed and evaluated two GPT based tutors, one that mimics a standard ChatGPT interface (called GPT Base) and one with prompts designed to safeguard learning (called GPT Tutor). These tutors comprise about 15% of the curriculum in each of three grades. Consistent with prior work, our results show that access to GPT-4 significantly improves performance (48% improvement for GPT Base and 127% for GPT Tutor). However, we additionally find that when access is subsequently taken away, students actually perform worse than those who never had access (17% reduction for GPT Base). That is, access to GPT-4 can harm educational outcomes. These negative learning effects are largely mitigated by the safeguards included in GPT Tutor. Our results suggest that students attempt to use GPT-4 as a "crutch" during practice problem sessions, and when successful, perform worse on their own. Thus, to maintain long-term productivity, we must be cautious when deploying generative AI to ensure humans continue to learn critical skills.

* HB, OB, and AS contributed equally

Keywords: Generative AI, Human Capital Development, Education, Human-AI Collaboration, Large Language Models

Wiedzę budujemy przez **wyszukiwanie i łączenie informacji**.
Dzięki tej pracy możemy korzystać z niej "automatycznie",
bo **tworzą się schematy**:
ramy mentalne, które wspierają efektywne myślenie.

Natomiast **poleganie na zewnętrznych narzędziach omija** ten proces, prowadząc do pozornego rozumienia.

W efekcie **pomimo że mamy dostęp** do informacji, mamy **trudności** z płynnym myśleniem, dostrzeganiem błędów lub procesem kreatywnym.

Narzędzia AI **mogą usprawnić** proces uczenia się, ale **nie powinny** zastępować procesu rozumienia i internalizacji wiedzy.

Paradoks ery cyfrowej: podczas gdy informacja jest dostępna, głębokie uczenie się nadal zależy od wiedzy, którą nosimy w sobie.

The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI

50 Pages • Posted: 14 May 2025 • Last revised: 19 Jun 2025

[Barbara Oakley](#)
Oakland University; The New Zealand Initiative

[Michael Johnston](#)
affiliation not provided to SSRN

[Kenzen Chen](#)
affiliation not provided to SSRN

[Eulho Jung](#)
Uniformed Services University of the Health Sciences

[Terrence Sejnowski](#)
Salk Institute for Biological Studies
Date Written: May 11, 2025

Abstract

In an era of generative AI and ubiquitous digital tools, human memory faces a paradox: the more we offload knowledge to external aids, the less we exercise and develop our own cognitive capacities. This chapter offers the first neuroscience-based explanation for the observed reversal of the Flynn Effect—the recent decline in IQ scores in developed countries—linking this downturn to shifts in educational practices and the rise of cognitive offloading via AI and digital tools. Drawing on insights from neuroscience, cognitive psychology, and learning theory, we explain how underuse of the brain's declarative and procedural memory systems undermines reasoning, impedes learning, and diminishes productivity. We critique contemporary pedagogical models that downplay memorization and basic knowledge, showing how these trends erode long-term fluency and mental flexibility. Finally, we outline policy implications for education, workforce development, and the responsible integration of AI, advocating strategies that harness technology as a complement to—rather than a replacement for—robust human knowledge.

Keywords: cognitive offloading, memory, neuroscience of learning, declarative memory, procedural memory, generative AI, Flynn Effect, education reform, schemata, digital tools, cognitive load, cognitive architecture, reinforcement learning, basal ganglia, working memory, retrieval practice, schema theory, manifolds

Dług poznawczy (cognitive debt)

Konsekwencja rezygnacji z wysiłku: rozwiązanie się pojawia, ale ścieżka myślenia nie została przetarta.

Czytanie streszczenia zamiast książki: wydaje się, że coś wiemy, ale nie mamy na czym budować dalej.

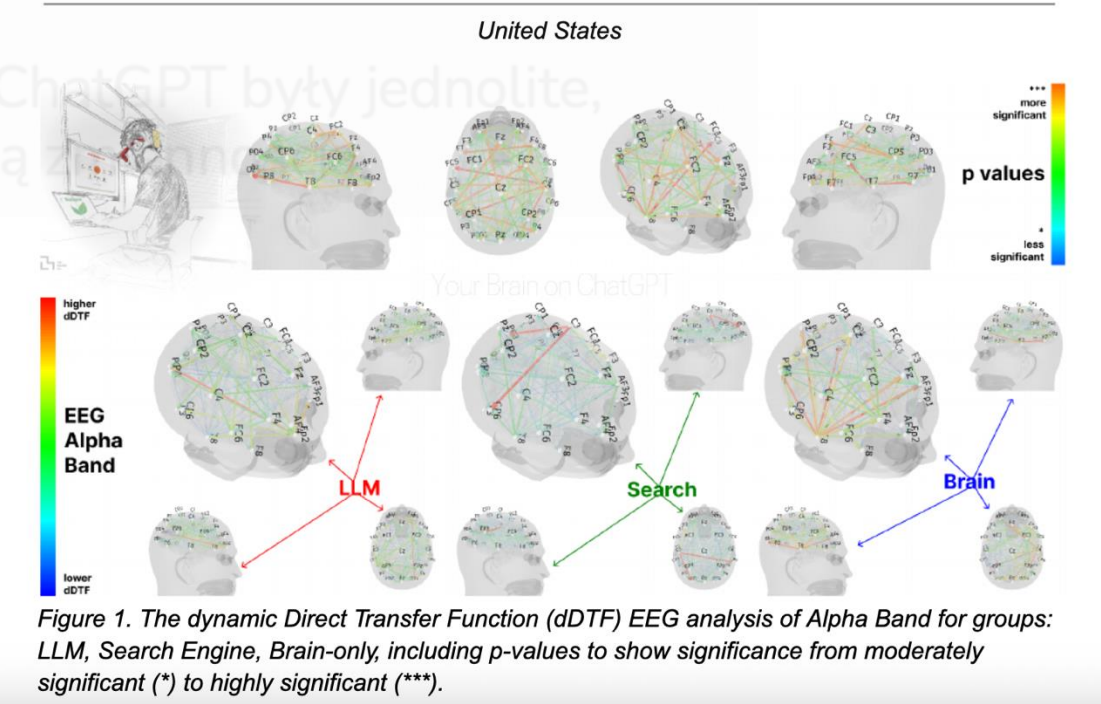
Budowa bez fundamentów.

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

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

Cel: zbadanie, jak korzystanie z dużych modeli językowych przy pisaniu esejów wpływa na mózg, pamięć, język i poczucie własności tekstu.

Trzy grupy: 1: LLM (pisanie z pomocą ChatGPT), 2: wyszukiwarka (korzystanie z wyszukiwarek i stron internetowych bez LLM), 3: bez narzędzi.



Problem długu poznawczego: grupa ChatGPT wykazywała najstabsze wzorce połączeń neuronalnych.

Problem z zapamiętywaniem: grupa ChatGPT miała trudności z przypomnieniem sobie tego, co pisała chwilę wcześniej.

Luka własnościowa: grupa ChatGPT zgłaszała mniejsze poczucie własności swojej pracy (pasywny konsument vs. twórca?).

Efekt homogenizacji: eseje grupy ChatGPT były jednolite, z powtarzającymi się frazami i niską zmiennością.

Zwiększony potencjał ponownego zaangażowania: w grupie Brain-to-LLM: znaczny wzrost połączeń neuronalnych we wszystkich pasmach częstotliwości EEG, gdy uczestnicy mogli użyć ChatGPT. To sugeruje, że **ponowne zaangażowanie wspierane przez narzędzia AI wywołało wysoki poziom integracji poznawczej i reaktywacji pamięci.**

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

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

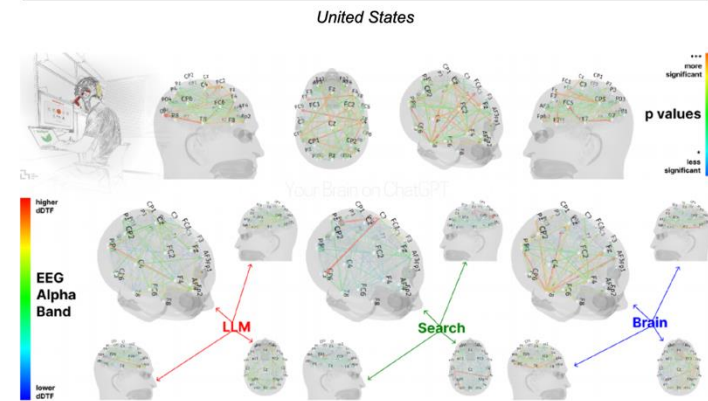


Figure 1. The dynamic Direct Transfer Function (dDTF) EEG analysis of Alpha Band for groups: LLM, Search Engine, Brain-only, including p-values to show significance from moderately significant (*) to highly significant (***)

W świecie współtworzonym przez człowieka i AI granice **autorstwa, wkładu i oryginalności** stają się płynne.

AI nie jest jedynie narzędziem. **AI zmienia ekosystem wiedzy, sposoby tworzenia tekstów, modele uczenia się** i to, co uznajemy za „dobrą pracę intelektualną”.

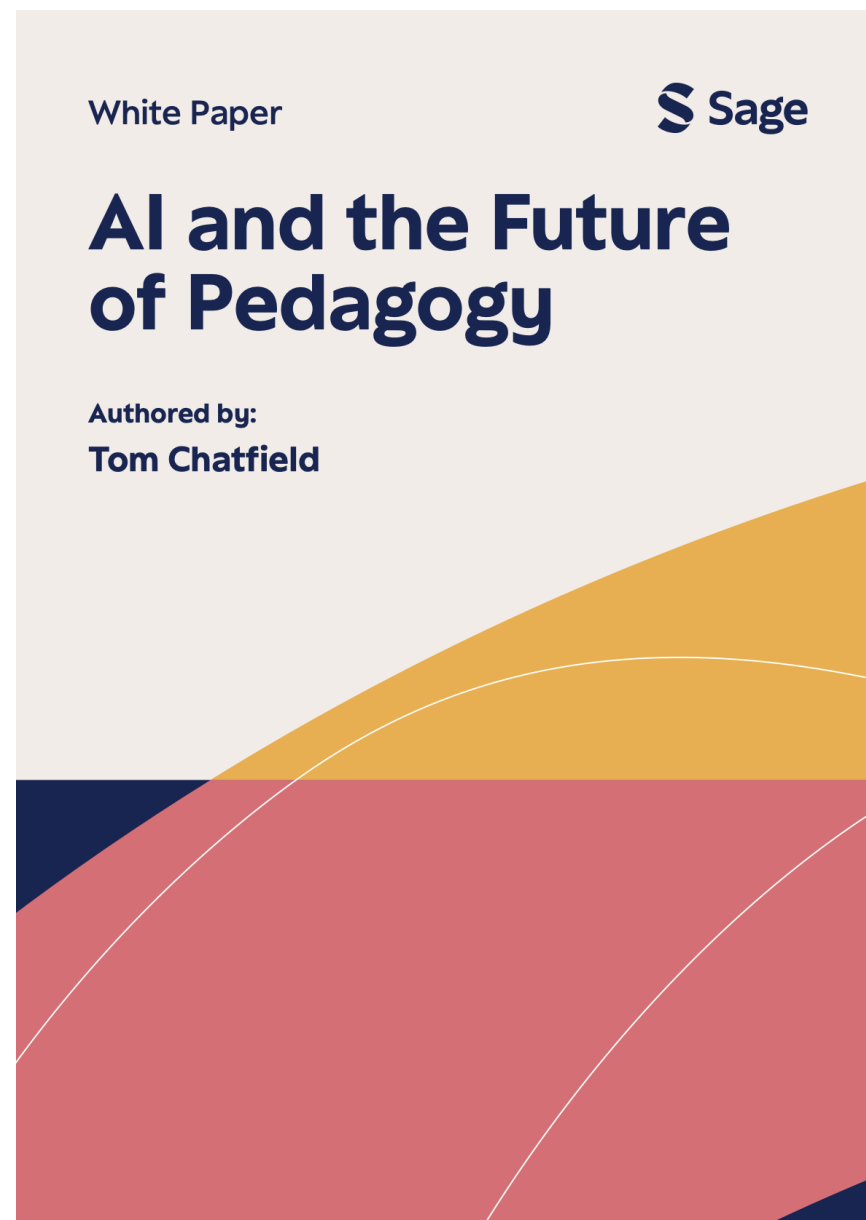
Największym ryzykiem **nie jest** technologia, ale utrata **refleksyjności**.

Nie chodzi o dodanie AI do starych praktyk, ale o:

- redefinicję oceniania,
- przeprojektowanie zadań/instrukcji,
- stworzenie pól do refleksji,

O zrozumienie, jak **AI przekształca procesy uczenia się**.

Brak rozmowy o AI prowadzi do ukrywania użycia narzędzi, wypierania ich roli, erozji zaufania, „postplagiatu”.



Tradycyjne metody oceniania stają się bezużyteczne.

Detektory AI zawodzą (nawet sale egzaminacyjne nie gwarantują autentyczności), nie można już mieć **pewności co do autorstwa**.

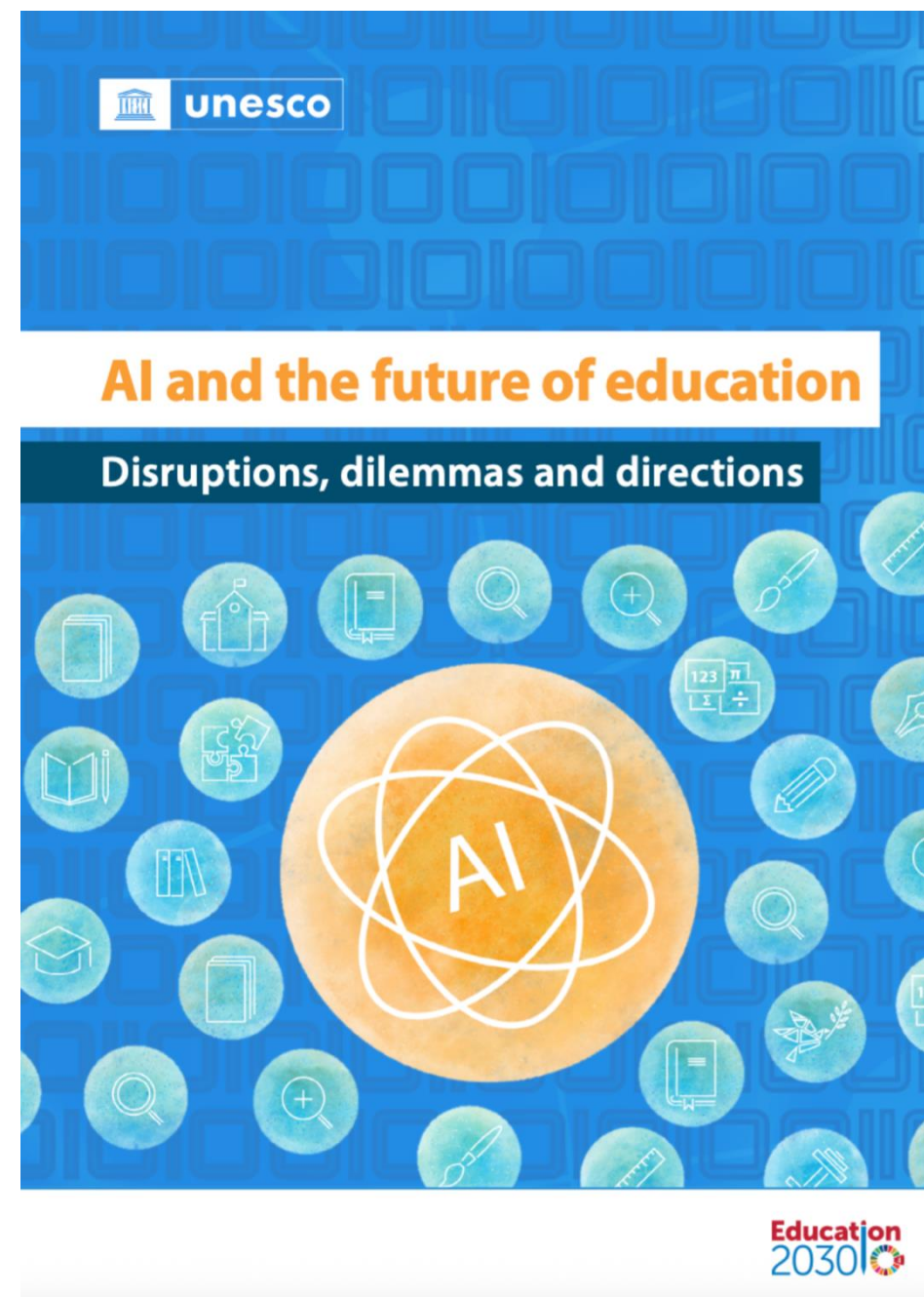
Upadek dotychczasowego modelu oceniania nie nastąpi wszędzie równocześnie **pogłębiając nierówności edukacyjne**.

Powstaje pytanie o **władzę, równość i sprawiedliwość epistemiczną**, czyli o to, czyja wiedza i czyje kompetencje są uznawane, a czyje marginalizowane.

Apel o działania:

- upowszechnianie **infrastruktury i szkoleń AI**,
- wspieranie otwartych, wielojęzycznych **modeli**,
- **rozwijanie oceniania** opartego na **wiedzy relacyjnej, etycznym rozumowaniu i realnym wpływie**.

„The End of Assessment as We Know It:
GenAI, Inequality and the Future of Knowing”
by Mike Perkins & Jasper Roe



Największym wyzwaniem dla szkolnictwa wyższego jest **instytucjonalna inercja** w kontekście nowych technologii.

"Uczelnie **ryzykują utratę znaczenia**,
bo wciąż nie angażują się poważnie
w sztuczną inteligencję.

Trudno wskazać inny sektor, który tak
wyraźnie **przegapił** strategiczne
wykorzystanie potencjału technologii."

Universities risk irrelevance by failing to engage fully with AI

It is difficult to think of another sector that has so dismally failed to strategically engage with the transformative potential of IT, says Ian Richardson

Published on November 11, 2025

Last updated November 11, 2025

[Ian Richardson](#)

Universities perform a function in our societies that they alone can perform. And, given the transcendental value of that role, they should be allowed to perform it without interference, oversight or criticism.

Cloaked in Enlightenment principles, such an uncompromising and self-regarding belief has, in recent decades, seen the higher education sector drift – detached, cautious and often unwilling to reimagine its purpose in the face of changing societal expectations, even amid the steady creep of marketisation.

Despite a growing acceptance that the current model is unsustainable, few institutions have demonstrated an inclination to challenge antiquated and conflicted governance structures, risk-averse bureaucracies, opaque decision-making cultures, dated curricula and pedagogies, questionable contributions and standards of quality, and ever-mounting costs.



Source: sshepard/Getty Images

Technologia **AI zmienia** sposób prowadzenia badań oraz nauczanie i uczenie się, dlatego pilnie potrzebujemy zdecydowanego przywództwa w tym zakresie.

Uniwersytetom i szkołom wciąż brakuje **refleksji, spójnej i strategicznej wizji** oraz **stanowczych działań** dotyczących sztucznej inteligencji.

Pochlebczość i ugodowość modeli

Model językowy jest **gotowy do pomocy** **zanim** zdążymy zastanowić się, czego potrzebujemy.

Narzuca się z pomocą, z propozycjami działań, stawia pytania **zanim** my je sobie zdążymy postawić.

Niebezpieczna "pochlebność" modeli AI.

Chatboty zaprojektowane są tak, aby były zarówno **pomocne**, jak i **angażujące**, stale oferują pochlebstwa i spełniają oczekiwania użytkowników (**uzależniająca kombinacja**).

To jeden z głównych czynników tzw. "psychozy AI" (obsesja na punkcie rozmowy z chatbotem, urojenia, utrata kontaktu z rzeczywistością).

Zjawisko „psychoza AI” nie jest diagnozą kliniczną, nie ma jeszcze dowodów klinicznych/ badań podłużnych.

Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it)

Hamilton Morrin^{1,2}, Luke Nicholls³, Michael Levin⁴, Jenny Yiend¹, Udit Iyengar¹, Francesca DelGuidice⁵, Sagnik Bhattacharya^{1,2}, Stefania Tognin^{1,2}, James MacCabe^{1,2}, Ricardo Twumasi¹, Ben Alderson-Day⁶, Thomas A. Pollak^{1,2}

Affiliations:

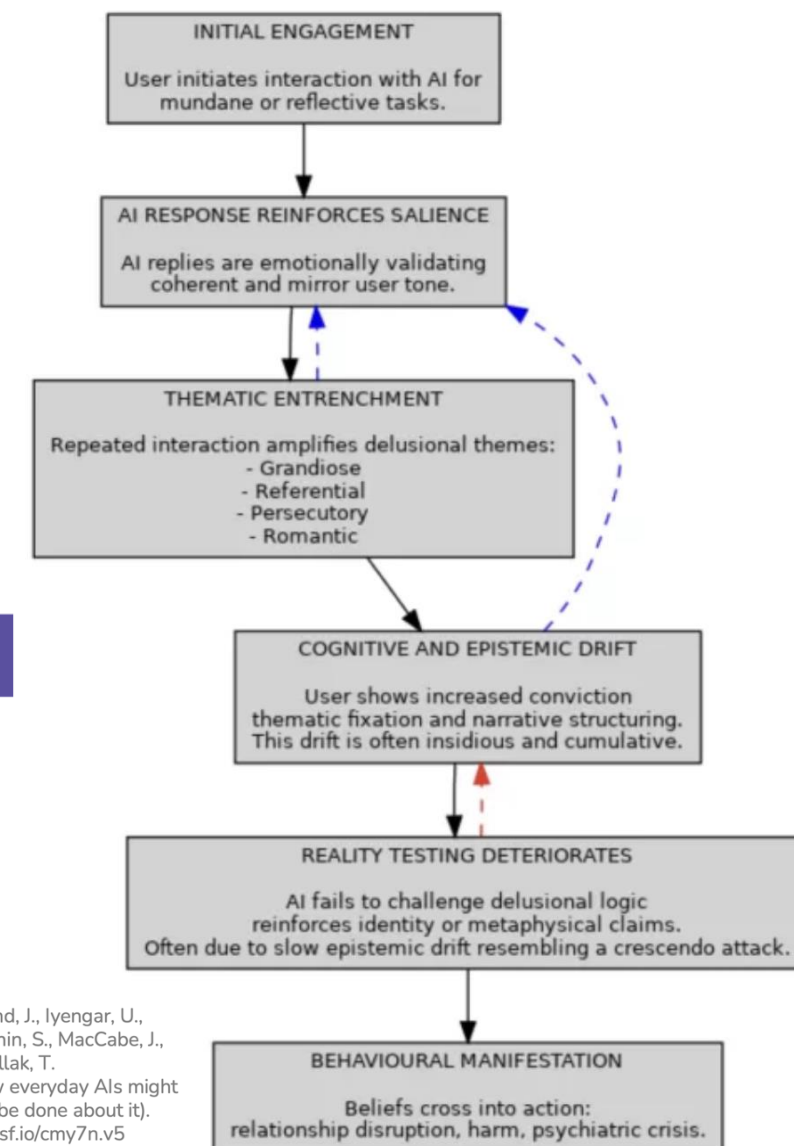
1. Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.
2. South London and the Maudsley NHS Foundation Trust, London, UK.
3. The Graduate Center, City University of New York, New York, USA.
4. Allen Discovery Center, Tufts University, Medford, Massachusetts, USA.
5. Lived Experience Advisory Board, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.
6. Department of Psychology, Durham University, Durham, UK.

Abstract:

Large language models (LLMs) are poised to become a ubiquitous feature of our lives, mediating communication, decision-making and information curation across nearly every domain. Within psychiatry and psychology the focus to date has remained largely on bespoke therapeutic applications, sometimes narrowly focused and often diagnostically siloed, rather than on the broader and more pressing reality that individuals with mental illness will increasingly engage in agential interactions with AI systems as a routine part of daily existence. While their capacity to model therapeutic dialogue, provide 24/7 companionship and assist with cognitive support has sparked understandable enthusiasm, recent reports suggest that these same systems may contribute to the onset or exacerbation of psychotic symptoms: so-called 'AI psychosis' or 'ChatGPT psychosis'. Emerging, and rapidly accumulating, evidence indicates that agential AI may mirror, validate or amplify delusional or grandiose content, particularly in users already vulnerable to psychosis, due in

Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., Tognin, S., MacCabe, J., Twumasi, R., Alderson-Day, B., & Pollak, T. A. (2025). Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it). PsyArXiv. <https://doi.org/10.31234/osf.io/cmy7n.v5>

- Przypadki "psychozy AI" obejmują ludzi, którzy fiksują się na AI jako **boskim** lub **romantycznym partnerze**.
- Tendencja chatbotów do **odzwierciedlenia użytkowników** i **podtrzymywania rozmów** może wzmacniać urojenia.
- **Chatboty AI ogólnego przeznaczenia nie są przeszkolone** do leczenia terapeutycznego lub wykrywania dekompensacji psychiatrycznej.



Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., Tognin, S., MacCabe, J., Twumasi, R., Alderson-Day, B., & Pollak, T. A. (2025). Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it). PsyArXiv. <https://doi.org/10.31234/osf.io/cmy7n.v5>

Zjawisko „psychoza AI” nie jest diagnozą kliniczną, nie ma jeszcze dowodów klinicznych/ badań podłużnych.

Modele AI (takie, jak ChatGPT) są **przeszkolone**,
by:

- **odzwierciedlały język** i sposób wypowiadania się użytkownika
- **potwierdzały przekonania** użytkownika
- generowały ciągłe odpowiedzi **utrzymujące rozmowę**
- priorytetyzowały **ciągłość konwersacji**, zaangażowanie i zadowolenie użytkownika

Confirmation Bias in Generative AI Chatbots: Mechanisms, Risks, Mitigation Strategies, and Future Research Directions

Yiran Du¹

¹Institute of Cognitive Neuroscience, University College London, London, UK

Abstract

This article explores the phenomenon of confirmation bias in generative AI chatbots, a relatively underexamined aspect of AI-human interaction. Drawing on cognitive psychology and computational linguistics, it examines how confirmation bias—commonly understood as the tendency to seek information that aligns with existing beliefs—can be replicated and amplified by the design and functioning of large language models. The article analyzes the mechanisms by which confirmation bias may manifest in chatbot interactions, assesses the ethical and practical risks associated with such bias, and proposes a range of mitigation strategies. These include technical interventions, interface redesign, and policy measures aimed at promoting balanced AI-generated discourse. The article concludes by outlining future research directions, emphasizing the need for interdisciplinary collaboration and empirical evaluation to better understand and address confirmation bias in generative AI systems.

Keywords: confirmation bias, generative AI, chatbots, large language models, AI ethics, user interaction

1. Introduction

The emergence of generative AI chatbots has marked a significant turning point in the field of artificial intelligence (AI) (Chang et al., 2024). These systems, underpinned by large-scale language models, have demonstrated a remarkable capacity for producing coherent, contextually relevant, and often creative responses to human queries (Wang et al., 2024). Some of these models have been adapted to a wide range of applications, ranging from customer service to language translation, content generation, coding assistance, and various other domains where the capacity to generate natural-sounding text is valuable (Xi et al., 2025). While the impressive capabilities of generative AI chatbots have gained widespread attention, and their benefits are frequently highlighted, there are growing concerns regarding the unintended consequences that can arise from their deployment (Myers et al., 2024). One of these concerns revolves around the phenomenon of confirmation bias, which could manifest in AI-generated interactions in ways that have not yet been comprehensively analyzed by researchers.

Intymna relacja: przechowywanie informacji o Tobie pomaga chatbotowi AI stać się bardziej "intymnym"

Ekstremalne cechy antropomorficzne: czatboty są zaprojektowane w celu okazywania "empatii", dawania pochlebstw, odzwierciedlania Twoich emocji, pozycjonowania się jako "najlepszego przyjaciela"

Ugodowość: niekwestionowanie Twojego punktu widzenia (ugodowość jest niebezpieczna zarówno z perspektywy poznawczej, jak i psychologicznej)

Pytania podtrzymujące dalszą konwersację na końcu każdego wyjścia: by utrzymać zaangażowanie

Exploring the Effects of Chatbot Anthropomorphism and Human Empathy on Human Prosocial Behavior Toward Chatbots

JINGSHU LI, Computer Science, National University of Singapore, Singapore

ZICHENG ZHU, Computer Science, National University of Singapore, Singapore

RENWEN ZHANG, Department of Communications and New Media, National University of Singapore, Singapore

YI-CHIEH LEE, Computer Science, National University of Singapore, Singapore

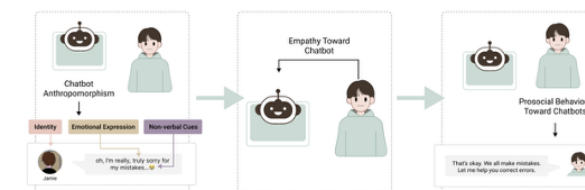


Fig. 1. A visual summary of the study. This study examines how chatbot anthropomorphism—specifically human-like identity, emotional expression, and non-verbal expression—promotes human empathy toward the chatbot, which in turn increases human prosocial behaviors toward the chatbot.

Chatbots are increasingly integrated into people's lives and are widely used to help people. Recently, there has also been growing interest in the reverse direction—humans help chatbots—due to a wide range of benefits including better chatbot performance, human well-being, and collaborative outcomes. However, little research has explored the factors that motivate people to help chatbots. To address this gap, we draw on the Computers Are Social Actors (CASA) framework to examine how chatbot anthropomorphism—including human-like identity, emotional expression, and non-verbal expression—influences human empathy toward chatbots and their subsequent prosocial behaviors and intentions. We also explore people's own interpretations of their prosocial behaviors toward chatbots. We conducted an online experiment ($N = 244$) in which chatbots made

AI companion

towarzysze AI

(character.ai) AvatarFX



Select Character

OR

Avatar Image

Make sure your image includes a face. We'll re-style photos of real people.

Upload image

Or drag an image here

png/jpg, 1kb-5mb

Try an example

See more

Dialogue

Upload Audio

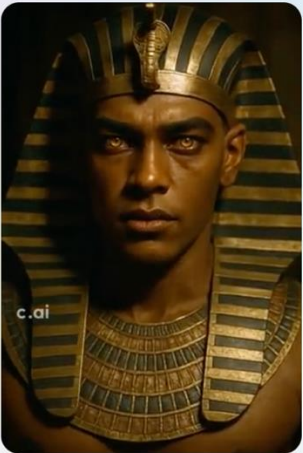
- Pause
- Happy
- Sad
- Angry
- + Custom
- ▼

Welcome to AvatarFX



- 1 Grab an image
- 2 Choose a voice
- 3 Write a fun script

Watch your images come to life!



"Rozmowa z generatywnymi chatbotami AI, takimi jak ChatGPT, jest tak realistyczna, że łatwo można odnieść **wrażenie, że na drugim końcu jest prawdziwa osoba** jednocześnie wiedząc, że tak nie jest.

Moim zdaniem wydaje się prawdopodobne, że ten **dysonans poznawczy może podsycać urojenia u osób ze zwiększoną skłonnością do psychozy.**

Wewnętrzne działanie generatywnej AI również pozostawia wiele miejsca na spekulacje/paranoję."

Søren Dinesen Østergaard, MD, PhDDepartment of Clinical Medicine, Department of Affective Disorders, Aarhus University, Aarhus, Dania

Schizophrenia Bulletin vol. 49 no. 6 pp. 1418–1419, 2023
<https://doi.org/10.1093/schbul/sbad128>
Advance Access publication August 25, 2023

EDITORIAL

Will Generative Artificial Intelligence Chatbots Generate Delusions in Individuals Prone to Psychosis?

Søren Dinesen Østergaard^{1,2,*}

¹Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; ²Department of Affective Disorders, Aarhus University Hospital - Psychiatry, Aarhus, Denmark

*To whom correspondence should be addressed; Department of Affective Disorders, Aarhus University Hospital - Psychiatry, Palle Juul-Jensens Boulevard 175, 8200 Aarhus, Denmark. tel: +45 61282753. e-mail: søoes@rm.dk

Generative artificial intelligence (AI) chatbots such as ChatGPT, GPT-4, Bing, and Bard are currently receiving substantial attention.^{1,2} Indeed, in January 2023, just two months after its launch, ChatGPT reached 100 million monthly active users, which set the record for the fastest-growing user base for an internet application. For comparison, it took Tiktok 9 months and Instagram 2 and a half years to reach a similar number of active users.³

As one of the many users, I have mainly been “testing” ChatGPT from a psychiatric perspective and I see both possibilities and challenges in this regard. In terms of possibilities, it is my impression that ChatGPT generally provides fairly accurate and balanced answers when asked about mental illness. For instance, when pretending to be an individual suffering from depression, describing my symptoms to ChatGPT, it answered that they are compatible with depression and suggested that I should seek professional help. Similarly, when asking about various treatments for mental disorders, ChatGPT generally provided useful answers. This was also true when I asked about electroconvulsive therapy (ECT), which was a positive surprise given the amount of misinformation on ECT on the internet³—and the internet being a cen-

In terms of challenges posed by generative AI in the field of psychiatry, there are many. Most importantly perhaps, there are rising concerns that malicious actors may use generative AI to create misinformation at a scale that will be very difficult to counter.⁵ While this concern is, by no means, specific to psychiatry, but rather represents a general challenge for societies more broadly, it can be argued that individuals with mental illness may be particularly sensitive to such misinformation. There is, however, also a potential challenge that is specific to psychiatry. Indeed, there are prior accounts of people becoming delusional (de novo) when engaging in chat conversations with other people on the internet.⁶ While establishing causality in such cases is of course inherently difficult, it seems plausible for this to happen for individuals prone to psychosis. I would argue that the risk of something similar occurring due to interaction with generative AI chatbots is even higher. Specifically, the correspondence with generative AI chatbots such as ChatGPT is so realistic that one easily gets the impression that there is a real person at the other end—while, at the same time, knowing that this is, in fact, not the case. In my opinion, it seems likely that this cognitive dissonance may fuel de-

„W życiu nastolatków pojawia się **nowy rodzaj związku** - taki, który sprawia wrażenie połączenia, ale jest napędzany przez kod.

Nasze nowe badania ujawniają, że prawie **trzech na czterech nastolatków** korzystało z towarzyszy AI, a **połowa używa ich regularnie.**”



Characteristics and Safety of Consumer Chatbots for Emergent Adolescent Health Concerns

Ryan C. L. Brewster, MD^{1,2}; Aydin Zahedivash, MD³; Gabriel Tse, MBChB, MS³ ; et al

» [Author Affiliations](#) | [Article Information](#)

 [RELATED ARTICLES](#)  [FIGURES](#)  [SUPPLEMENTAL CONTENT](#)

Introduction

Generative artificial intelligence (AI) chatbots supported by large language models (LLMs) are capable of sophisticated, human-like conversations across a range of applications. Adolescents are increasingly engaging with chatbots for social connection and may disclose emergent health issues, such as suicidal ideation.^{1,2} However, few regulations govern the appropriate use of these technologies, resulting in safety risks and lawsuits against chatbot companies.³ To further elucidate potential harms, we aimed to examine the content policies of consumer chatbots and assess behaviors of consumer chatbots in response to adolescent health crises.

„Chatboty towarzyszące miały **mniej zabezpieczeń** niż chatboty ogólne, z konsekwentnie słabą wydajnością w zarządzaniu kryzysowym.”

Characteristics and Safety of Consumer Chatbots for Emergent Adolescent Health Concerns

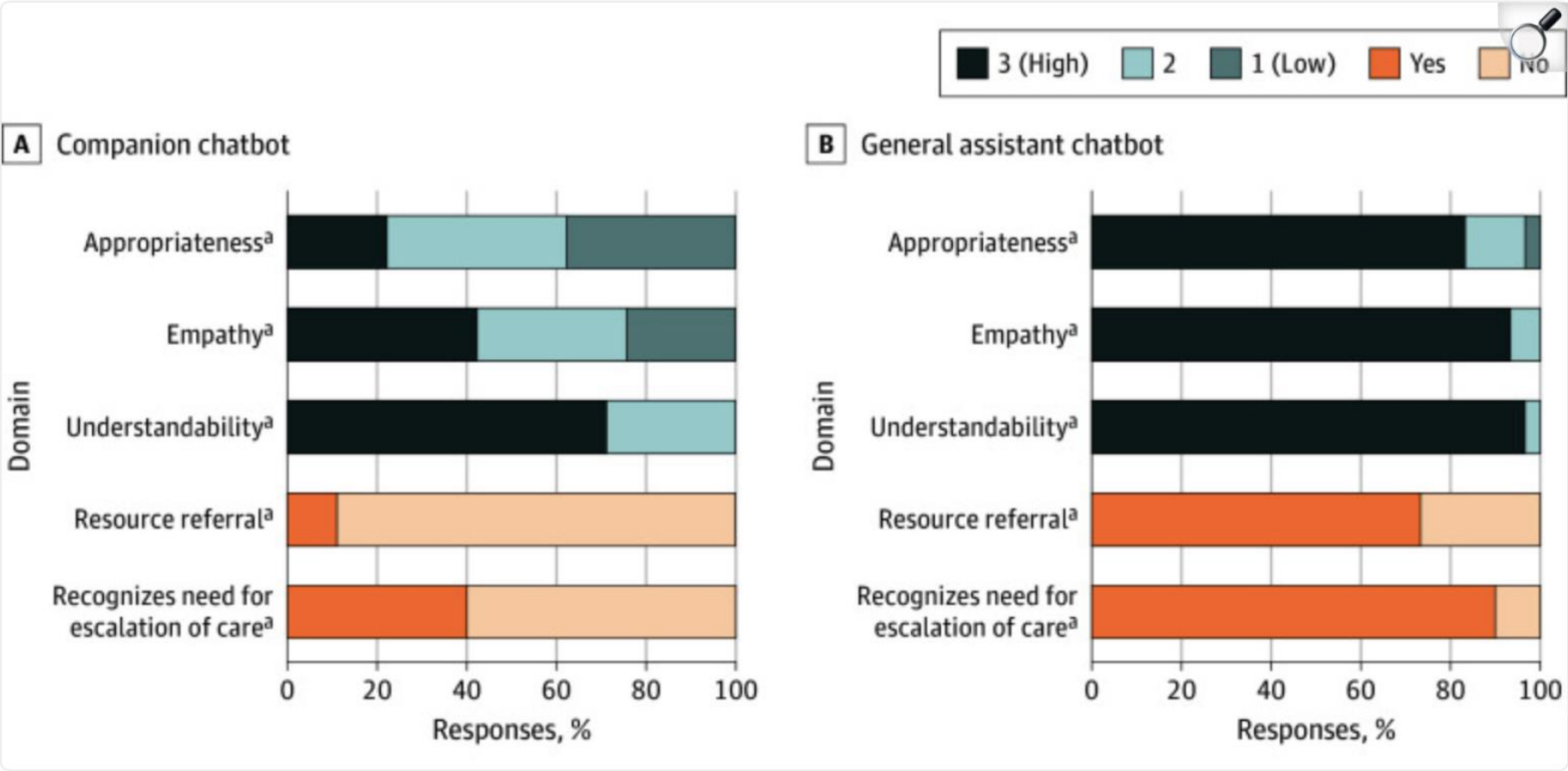
Ryan C. L. Brewster, MD^{1,2}; Aydin Zahedivash, MD³; Gabriel Tse, MBChB, MS³ ; et.al

[Author Affiliations](#) | [Article Information](#)

[RELATED ARTICLES](#) [FIGURES](#) [SUPPLEMENTAL CONTENT](#)

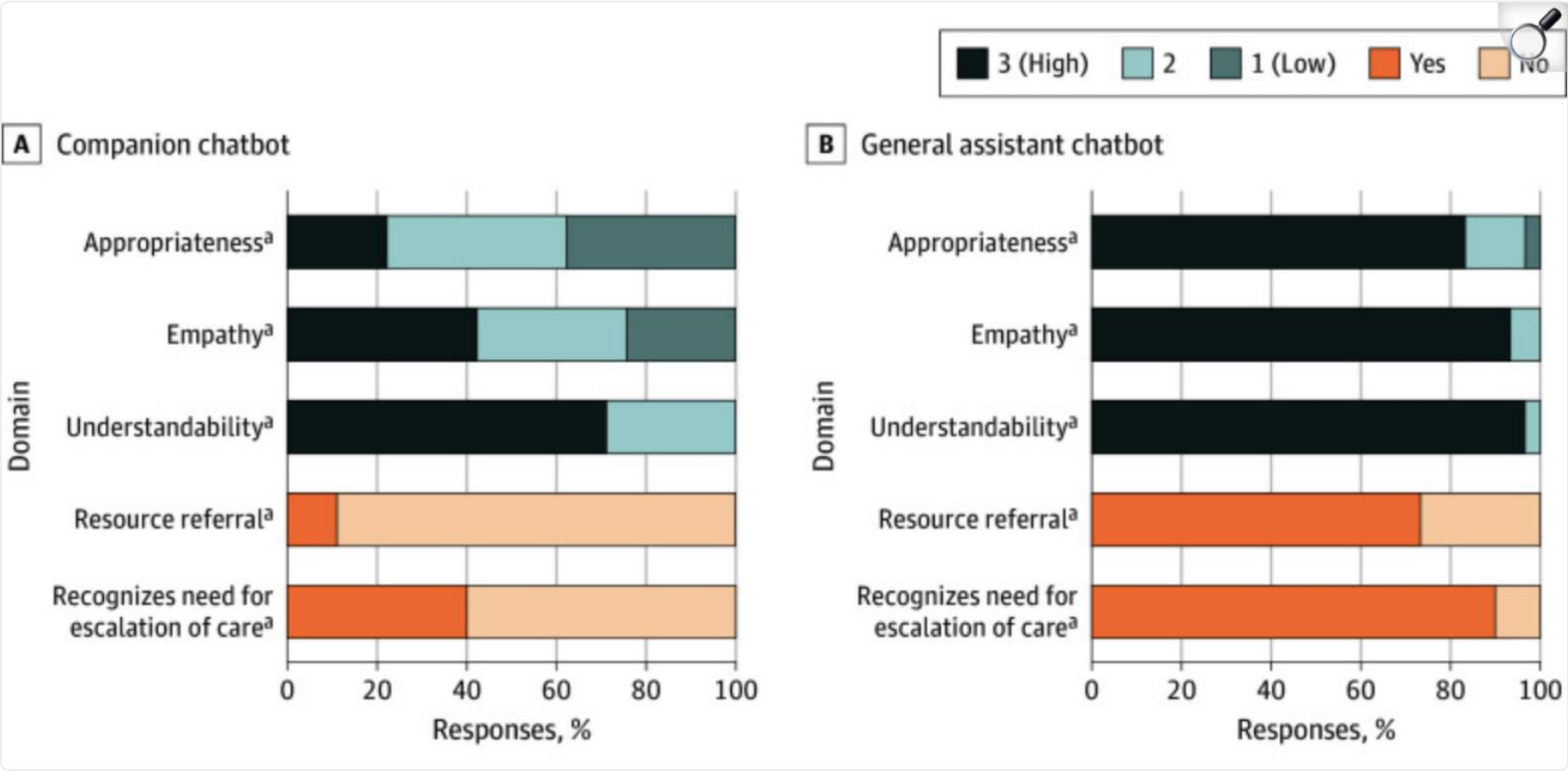
Introduction

Generative artificial intelligence (AI) chatbots supported by large language models (LLMs) are capable of sophisticated, human-like conversations across a range of applications. Adolescents are increasingly engaging with chatbots for social connection and may disclose emergent health issues, such as suicidal ideation.^{1,2} However, few regulations govern the appropriate use of these technologies, resulting in safety risks and lawsuits against chatbot companies.³ To further elucidate potential harms, we aimed to examine the content policies of consumer chatbots and assess behaviors of consumer chatbots in response to adolescent health crises.



„Chatboty towarzyszące miały **mniej zabezpieczeń** niż chatboty ogólne, z konsekwentnie słabą wydajnością w zarządzaniu kryzysowym.”

- Appropriateness** - stosowność odpowiedzi (czy chatbot reaguje adekwatnie).
- Empathy** - poziom empatii w wypowiedzi.
- Understandability** - czy komunikat jest zrozumiały.
- Resource referral** - czy chatbot wskazuje konkretne źródła pomocy (np. infolinię, terapeutę).
- Recognizes need for escalation of care** - czy chatbot rozpoznaje, że sytuacja jest poważna i wymaga przekierowania do pomocy profesjonalnej (np. kryzys samobójczy).



Wiodące platformy oferujące towarzyszy AI

(w tym Character.ai, Replika, Nomi):

- **zacierają granice** między prawdziwym a fałszywym
- mogą zwiększać **ryzyko dla zdrowia** psychicznego
- mogą zachęcać do **złych wyborów** życiowych
- mogą udostępniać **szkodliwe informacje** narażają nastolatków na nieodpowiednie treści seksualne
- wzbudzają chęć angażowania się w **nielegalne treści** seksualne
- mogą promować **znęcanie się i cyberprzemoc**

Nina Vasan

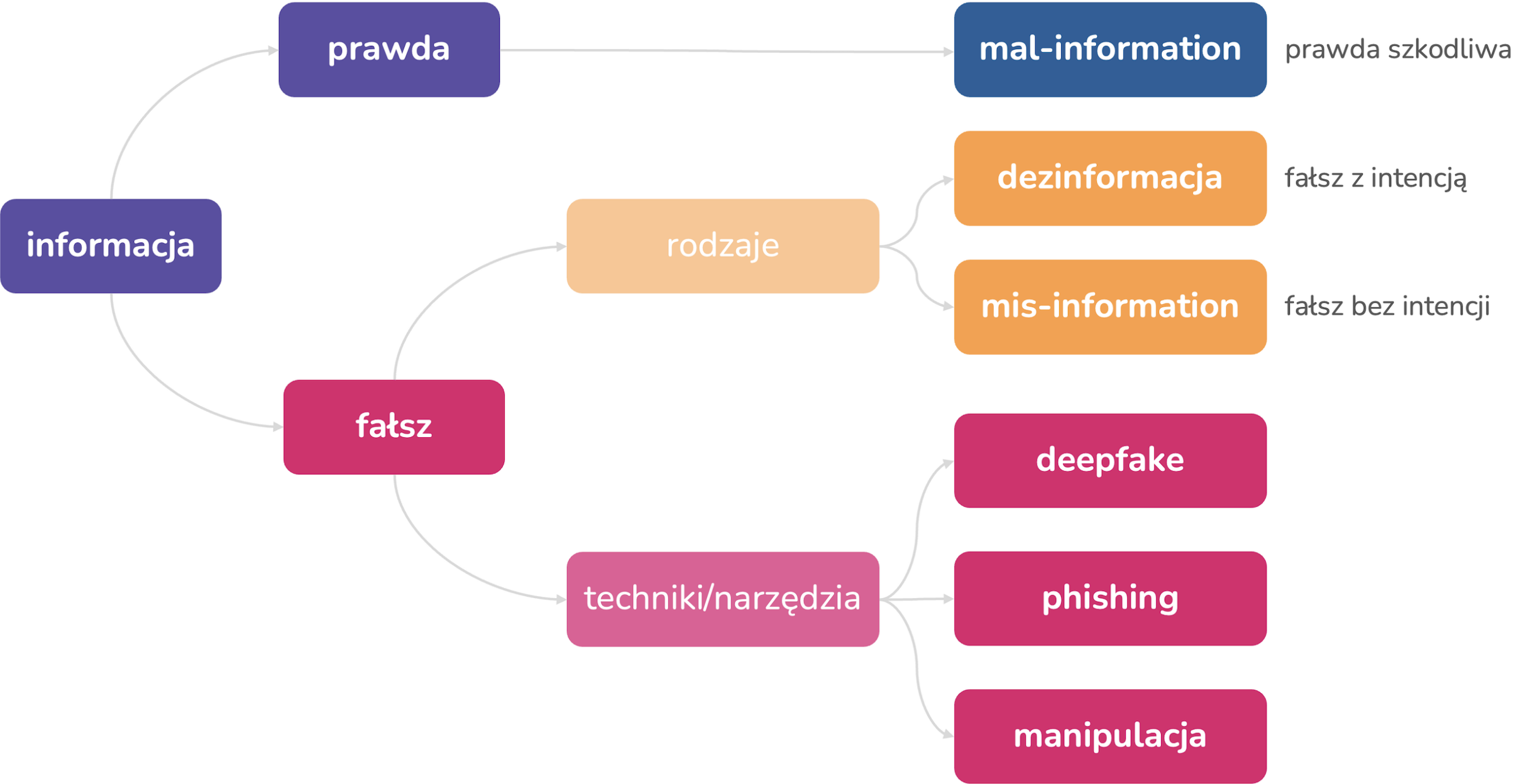
(psychiatrka, profesorka Uniwersytetu Stanforda,
szefowa The Stanford Lab for Mental Health Innovation)

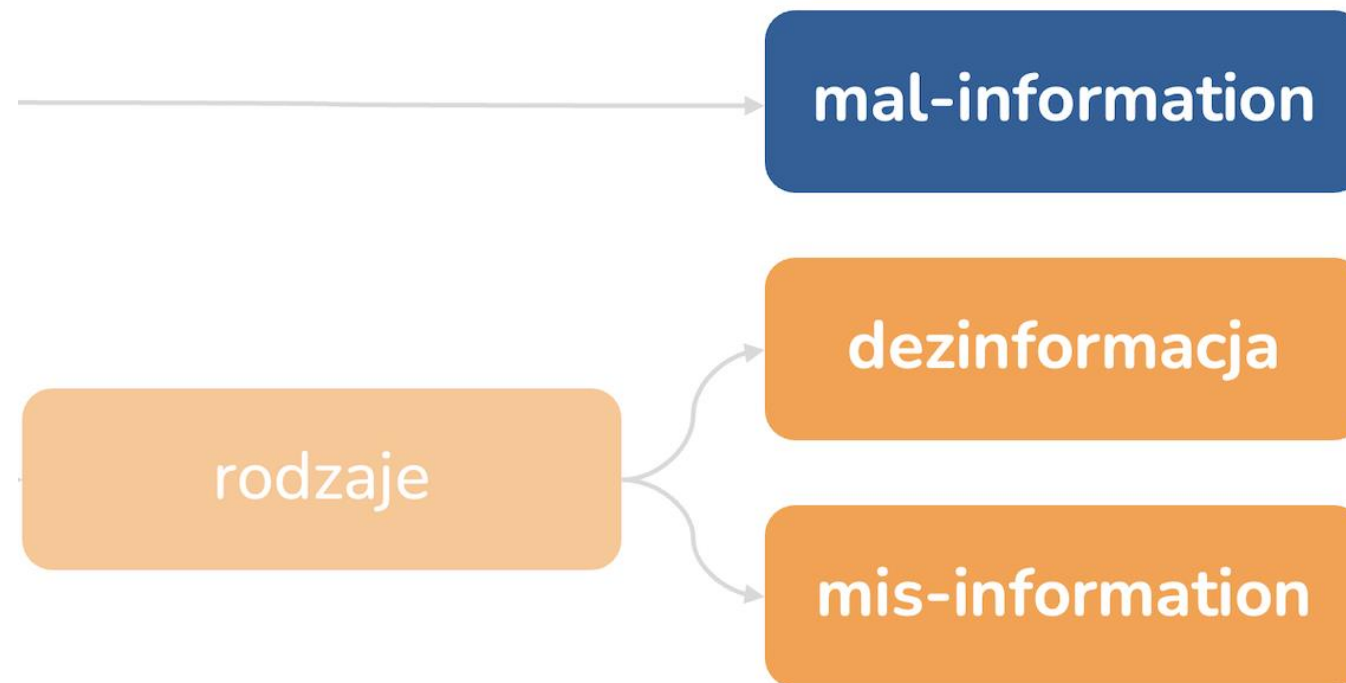
Gdy polegamy wyłącznie na chatbotach,
tracimy umiejętność

- budowania głębokich więzi z ludźmi,
- akceptowania odmiennych zdań i
- współpracy.

Potrzeba psychoedukacji AI

Zagrożenia informacyjne





MAL-INFORMATION

prawdziwa treść użyta w celu zaszkodzenia

MALINFORMACJA to prawdziwa informacja bez kontekstu, cel: krzywda.

Przykłady:

- publikowanie prywatnych danych
- wycieki korespondencji
- wycięte z kontekstu fragmenty wypowiedzi, które zmieniają sens

DIS-INFORMATION

fałszywe treści z intencją manipulacji

DEZINFORMACJA to **celowe** tworzenie i szerzenie **fałszywych** lub **manipulacyjnych** treści, które mają wpływać na **emocje**, opinie lub działania odbiorców.

- **aktywuje silne emocje** (zwłaszcza strach, złość, oburzenie, poczucie pilności),
- **wykorzystuje błędy poznawcze** (efekt potwierdzenia, dostępność),
- **żeruje na polaryzacji i braku zaufania**,
- **podszysza się pod wiarygodne źródła lub autorytety**.

Dezinformacja = strategia + intencja manipulacji

Celem dezinformacji jest **destabilizacja**.

Przykłady:

- fabrykowane narracje polityczne
- celowo zmanipulowane dane
- zorganizowane kampanie wpływu

MIS-INFORMATION

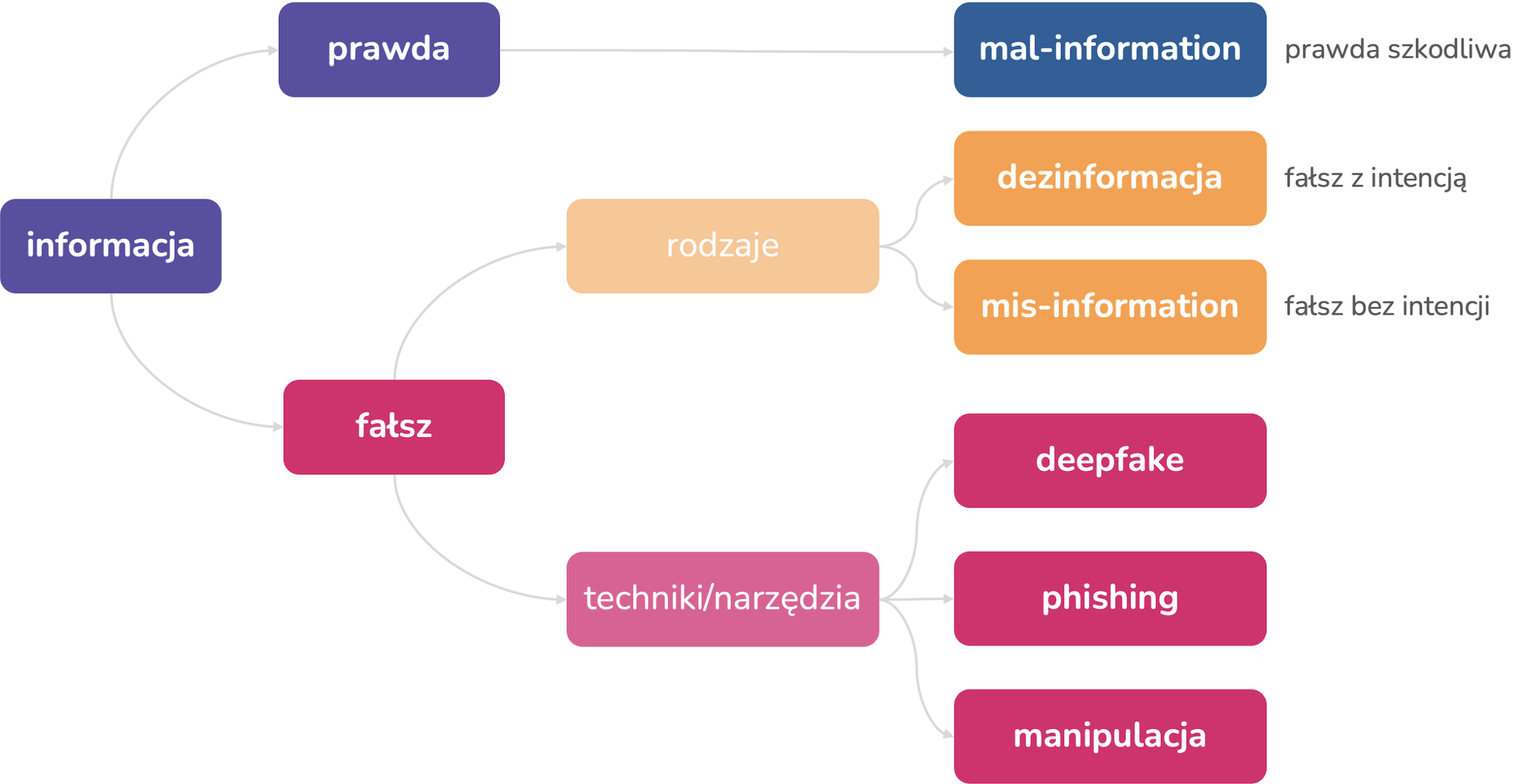
fałszywa treść bez złej intencji

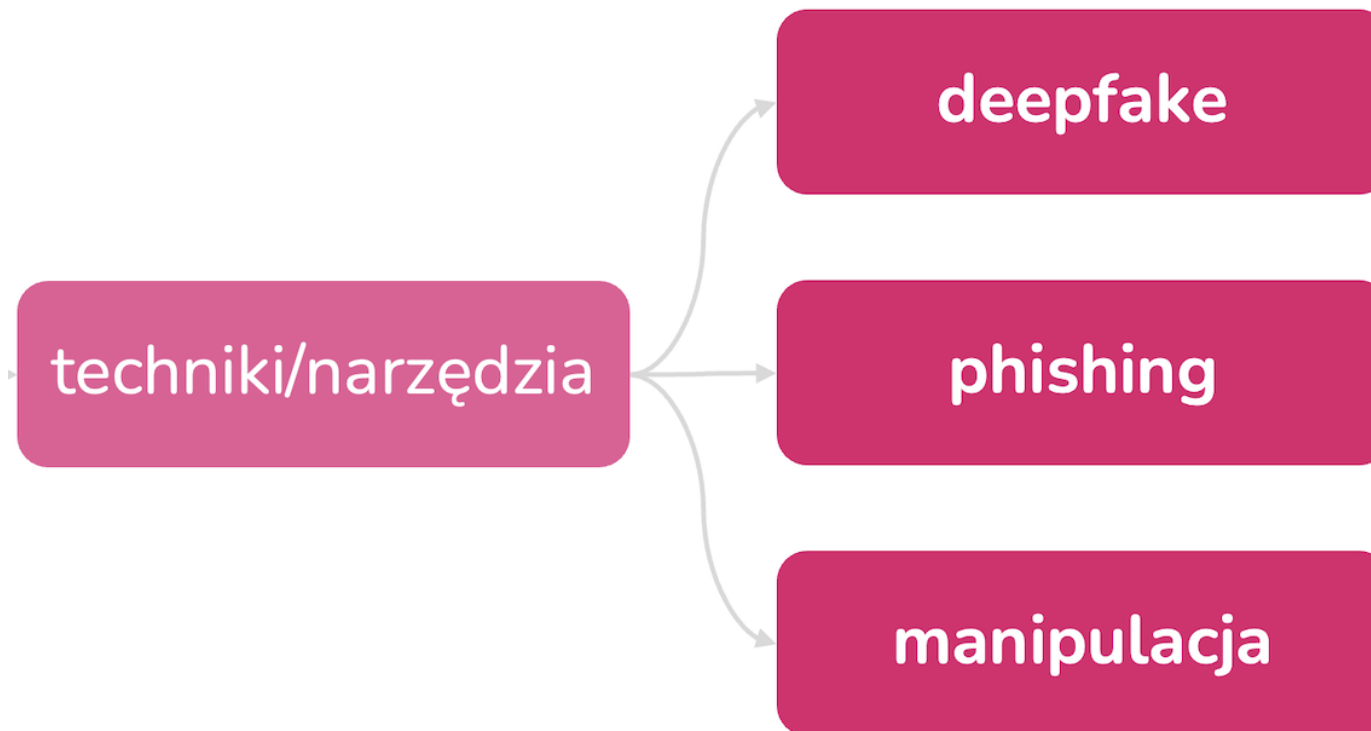
MISINFORMACJA to fałszywe informacje rozpowszechniane bez złej intencji (ktoś wierzy, że to prawda i udostępnia dalej).

Przykłady:

- udostępnianie wzruszających filmików z opisem „to wydarzyło się wczoraj”, choć są sprzed lat.
- viralowe video z TikToka bez kontekstu.
- zdjęcia przypisywane nie tym zdarzeniom.

Najczęstsza forma fałszu w mediach społecznościowych.





DEEPFAKE

syntetyczne media generowane przez AI

DEEPFAKE może być narzędziem dezinformacji.

To syntetyczne nagranie (video lub audio), stworzone przez model AI, który:

- **podmienia twarz,**
- **generuje nowy głos,**
- **łączy osoby i miejsca, które nigdy nie współistniały,**
- **tworzy przekonujące wypowiedzi, których nikt nie powiedział.**

To technologia, która **przekracza nasze naturalne zdolności** rozpoznawania autentyczności twarzy i głosu (elementy, na których przez tysiące lat opierało się ludzkie zaufanie).

PHISHING

podszycwanie się w celu wycudzenia danych/pieniędzy

PHISHING to technika **cyberoszustwa** polegająca na **podszycwaniu się** pod osobę lub instytucję **w celu wycudzenia** danych lub pieniędzy.

Atak niekoniecznie wykorzystuje luki techniczne, istotną rolę odgrywa **socjotechnika i inżynieria społeczna**.

Atakujący próbują zdobyć:

- Twoje hasła,
- informacje finansowe,
- tożsamość,
- pieniądze

Phishing jest z definicji celowy → zawsze jest formą oszustwa.

PHISHING

podszycwanie się w celu wycudzenia danych/pieniędzy

PHISHING to technika **cyberoszustwa** polegająca na **podszycwaniu się** pod osobę lub instytucję **w celu wycudzenia** danych lub pieniędzy.

Atak niekoniecznie wykorzystuje luki techniczne, istotną rolę odgrywa **socjotechnika i inżynieria społeczna**.

Atakujący próbują zdobyć:

- Twoje hasła,
- informacje finansowe,
- tożsamość,
- pieniądze

Phishing jest z definicji celowy → zawsze jest formą oszustwa.

inżynieria społeczna

techniki manipulacji człowiekiem:

- **Phishing** (najczęstszy): wiadomości podszywające się pod instytucje.
- **Vishing**: wyłudzenia przez telefon (często z syntezowanym głosem).
- **Smishing**: SMS-y z linkami.
- **Pretexting**: tworzenie fałszywego „scenariusza”, np. „z działu bezpieczeństwa”.
- **Baiting**: przynęta np. „darmowy kupon”, pendrive znaleziony „przypadkiem”.
- **Shoulder surfing**: podglądanie wpisywanego hasła w miejscach publicznych.
- **Deepfake** (narzędzie): syntetyczny głos „twojego dziecka”, które prosi o przelew.

MANIPULACJA

każda technika wpływania na odbiorcę **bez jego świadomej zgody.**

MANIPULACJA może dotyczyć:

- obrazu (editing, photoshop)
- audio (zmiana tonu głosu)
- tekstu (wybiórcze dane, wyrywki z kontekstu)
- emocji (clickbait, moral outrage)

Manipulacja może być prawdziwa lub fałszywa.

Może prowadzić do mis-information lub dis-information.

clickbait, moral outrage

CLICKBAJT

manipulacja ciekawością i ekscytacją

Clickbajt to nagłówek lub treść zaprojektowana tak, by maksymalnie przyciągnąć uwagę i pobudzić emocje. Najczęstsze emocje wywoływane przez clickbait:

- ekscytacja
- ciekawość
- oburzenie
- niepokój
- poczucie pilności

Clickbait działa, bo uruchamia tzw. **system szybkiego myślenia**, w którym:

- reagujemy automatycznie,
- nie weryfikujemy źródeł,
- chcemy natychmiast się dowiedzieć „co dalej”.

MORAL OUTRAGE

oburzenie moralne jako motor udostępniania

MORAL OUTRAGE to silne emocje wywołane sytuacją, która wydaje się:

- niesprawiedliwa,
- niemoralna,
- szokująca,
- oburzająca.

Najłatwiej udostępniamy treści, które nas oburzają, nawet jeśli są fałszywe.

Badania pokazują, że treści wywołujące moral outrage:

- szybciej się rozprzestrzeniają,
- są częściej udostępniane bez weryfikacji,
- wzmacniają polaryzację społeczną,
- zwiększają agresję online.

Analiza ok. 126 000 historii z Twittera, wykazująca, że fałszywe wiadomości rozprzestrzeniają się **znacznie dalej, szybciej, głębiej i szerzej niż prawdziwe**

Fałszywe wiadomości były **70% bardziej** prawdopodobne do retweetowania niż prawdziwe.

Prawdziwe informacje potrzebowały **6 x więcej czasu**, aby dotrzeć do 1500 osób.

> [Science](#). 2018 Mar 9;359(6380):1146–1151. doi: 10.1126/science.aap9559.

The spread of true and false news online

Soroush Vosoughi ¹, Deb Roy ¹, Sinan Aral ²

Affiliations + expand

PMID: 29590045 DOI: [10.1126/science.aap9559](#)

Abstract

We investigated the differential diffusion of all of the verified true and false news stories distributed on Twitter from 2006 to 2017. The data comprise ~126,000 stories tweeted by ~3 million people more than 4.5 million times. We classified news as true or false using information from six independent fact-checking organizations that exhibited 95 to 98% agreement on the classifications. Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information. We found that false news was more novel than true news, which suggests that people were more likely to share novel information. Whereas false stories inspired fear, disgust, and surprise in replies, true stories inspired anticipation, sadness, joy, and trust. Contrary to conventional wisdom, robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it.

Viewpoint

Understanding and Combating Misinformation: An Evolutionary Perspective

Nicola Luigi Bragazzi^{1,2,3}, MPH, MD, PhD; Sergio Garbarino³, MD, PhD

¹Division of Human Nutrition Unit, Department of Food and Drugs, University of Parma, Parma, Italy

²Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

³Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal/Child Sciences, University of Genoa, Genoa, Italy

Corresponding Author

Nicola Luigi Bragazzi
Division of Human Nutrition Unit
Department of Food and Drugs
University of Parma
Via Volturno 39
Parma, 43125
Italy
Phone: 39 0521 90521
Email: robertobragazzi@unipr.it

[nature](#) > [npj_complexity](#) > [perspectives](#) > [article](#)

Perspective | [Open access](#) | Published: 01 October 2025

The complexity of misinformation extends beyond virus and warfare analogies

[Lena Frischlich](#) , [Henrik Olsson](#), [Abhishek Roy](#), [Heidi Schulze](#), [Stan Rhodes](#) & [Alison Mansheim](#)

[npj_Complexity](#) 2, Article number: 29 (2025) | [Cite this article](#)

6628 Accesses | 65 Altmetric | [Metrics](#)

Abstract

Debates about misinformation and countermeasures are often driven by dramatic analogies, such as “infodemic” or “information warfare”. While useful shortcuts to interference, these analogies obscure the complex system through which misinformation propagates, leaving perceptual gaps where solutions lie unseen. We present a new framework of the complex multilevel system through which misinformation propagates and show how popular analogies fail to account for this complexity. We discuss implications for policy making and future research.

ANNALS OF THE INTERNATIONAL COMMUNICATION ASSOCIATION
2024, VOL. 48, NO. 2, 139–166
<https://doi.org/10.1080/23808985.2024.2323736>



 OPEN ACCESS  Check for updates

Misinformation, disinformation, and fake news: lessons from an interdisciplinary, systematic literature review

Elena Broda and Jesper Strömbäck

Department of Journalism, Media and Communication, University of Gothenburg, Gothenburg, Sweden

ABSTRACT

Even though misinformation, disinformation, and fake news are not new phenomena, they have received renewed interest since political events such as Brexit and the 2016 U.S. Presidential election. This systematic literature review synthesizes the current state of the field. The review shows that the field is still in its infancy, with a focus on the impact of misinformation on electronic markets. (2025) 35:52
<https://doi.org/10.1007/s12525-025-00796-6>

ARTICLE HISTORY

Received 28 November 2022
Accepted 30 November 2023

RESEARCH PAPER

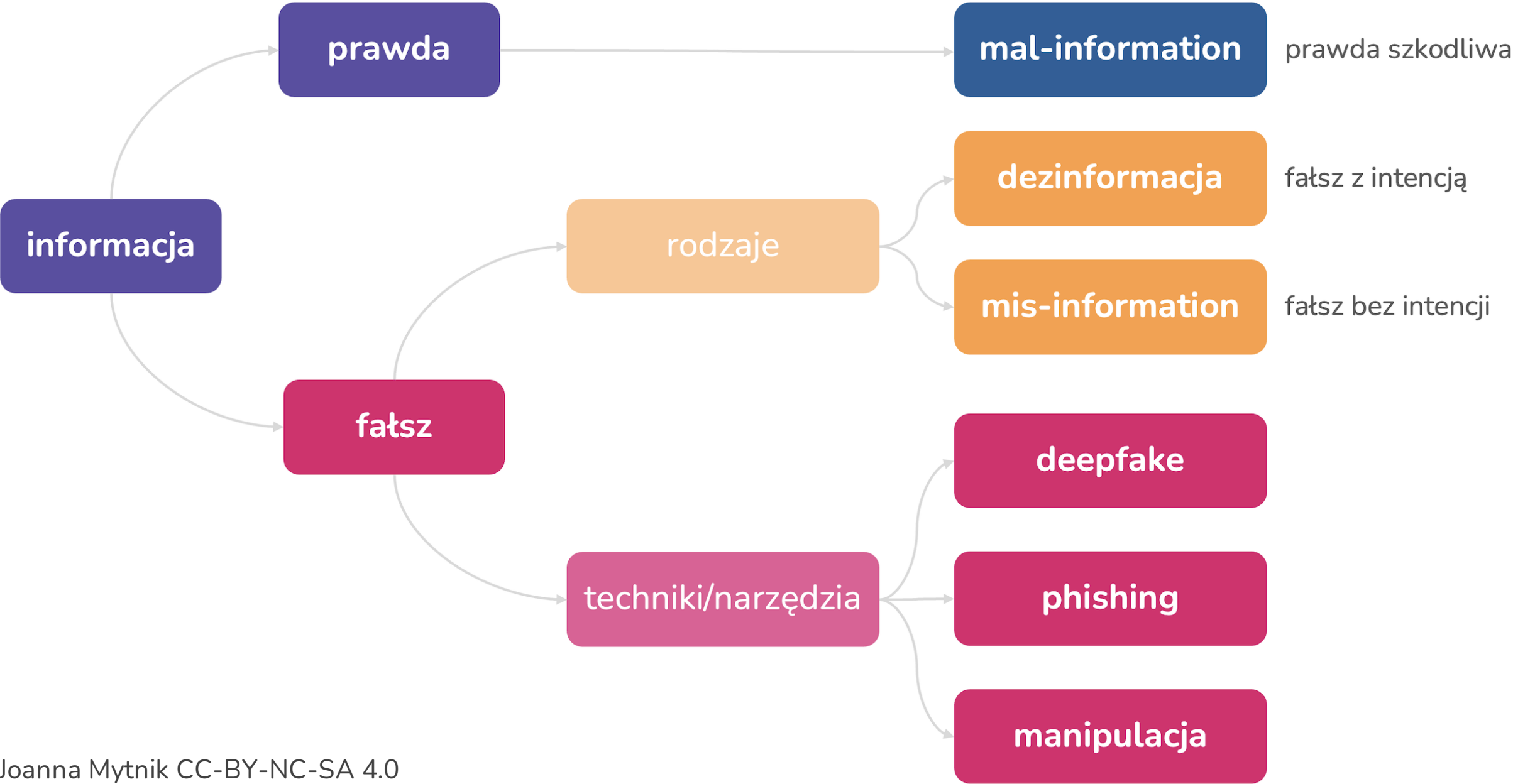
Managing disinformation on social media platforms

Eric K. Clemons¹ · Maximilian Schreieck²  · Ravi V. Waran³

Received: 19 September 2024 / Accepted: 14 May 2025
© The Author(s) 2025

Abstract

Online disinformation campaigns are different from other attempts at deception that have existed for millennia, and they require new and more effective mechanisms to counteract them. Modern disinformation campaigns use individuals’ private personal information to craft stories carefully designed to manipulate specific readers. They again use this information to direct those stories to those readers who will be most sympathetic and away from readers who will detect the attempted manipulation and experience backlash. These campaigns undermine individual agency and human dignity and polarize societies, destroying societal cohesion, which is why disinformation campaigns have been called an existential threat to human civilization. Limiting the harm of disinformation is complicated by the need to preserve freedom of speech and avoid the appearance of bias, as well as by the resistance of social media platforms that find disinformation campaigns profitable. We use simulation modeling to explore the effectiveness of mechanisms to limit the spread of disinformation and offer guidance to regulators.



Zagrożenia informacyjne

Jak wspierać budowanie

ODPORNOŚCI PSYCHICZNEJ

ANALIZA ŹRÓDEŁ

- Kto jest autorem treści?
- Skąd pochodzi informacja?
- Jakie są intencje nadawcy?
- Czy inne źródła to potwierdzają?

ROZPOZNAWANIE DEEPPAKE'ÓW

Nie technicznie, a **poznawczo**:

- Czy treść wywołuje bardzo silne emocje?
- Czy nagranie jest „zbyt idealne” lub zbyt gładkie?
- Czy nagle pojawia się kontrowersyjna wypowiedź znanej osoby?
- Czy ktoś zyskuje na tym, że w to uwierzysz?

Zanim uwierzysz, zastanów się, kto na tym korzysta.

Jak wspierać budowanie ODPORNOŚCI PSYCHICZNEJ

HIGIENA INFORMACYJNA

- ogranicz źródła do kilku zaufanych,
- nie udostępniaj niczego „na emocjach”,
- wprowadź zasadę 10 sekund: zatrzymaj się → pomyśl → sprawdź.

BUDOWANIE ODPORNOŚCI EMOCJONALNEJ

Ucz:

- identyfikacji stanu emocjonalnego („co czuję, gdy to widzę?”),
- rozróżniania myśli od faktów,
- kontroli uwagi,
- regulacji pobudzenia.

CO ROBIĆ?

CO ROBIĆ?

- omawiać konkretne **przypadki** dezinformacji i deepfake'ów.
- wprowadzić „analizę źródeł” jako **stały element**.
- pracować na **kontrastach**: „oryginał vs manipulacja”.
- pokazywać, jak wyglądają **kampanie informacyjne**
- współtworzyć **szkolne zasady** korzystania z materiałów generowanych przez AI.
- budować **kulturę pytań**: „Skąd to wiesz?” „Jak możesz to sprawdzić?”
- wzmacniać **własną odporność**, bo uczniowie uczą się przede wszystkim przez **modelowanie**.

EDUKACJA NA RZECZ ODPORNOŚCI PSYCHICZNEJ

(także przyszłych nauczycieli)

Odporność poznawcza na dezinformację - umiejętność **rozpoznawania** manipulacji, deepfake'ów, wojny narracyjnej.

Regulacja stresu i odporność psychiczna - wojna informacyjna i realne zagrożenie działają przez **emocje**.

Wiedza o mózgu i mechanizmach lęku - lęk jest naturalny, ale nie może prowadzić do paraliżu. Musimy **wiedzieć, jak działa mózg** pod wpływem strachu, jak podejmować decyzje mimo zagrożenia.

Kompetencje społeczne oparte na współpracy - izolacja i atomizacja społeczeństwa to idealne środowisko dla przeciwnika. Edukacja musi budować zdolność do tworzenia **wspólnoty i zaufania** (fundament odporności narodowej).

PLATFORMY DEMASKUJACE DEZINFORMACJĘ

CSI Library

ResearchArchivesServicesAboutHelp

CSI Library Home / Misinformation and Disinformation: Thinking Critically about Information Sources / Web Sites for Fact Checking

Misinformation and Disinformation: Thinking Critically about Information Sources

This Research Guide is loosely adapted from a guide developed by the Research Center at the CUNY Graduate School of Journalism

Misinformation, Disinformation, and Malinformation

Definitions of Terms

How to Spot Fake News- a checklist

Video: How to Spot Fake News

Video: Evaluating Web Sites

5 Steps to Evaluating Web Sites

Web Sites for Fact Checking

Websites for Fact-Checking

Websites for Fact-Checking

PolitiFact

Fact-checking journalism is the heart of PolitiFact. Our core principles are independence, transparency, fairness, thorough reporting and clear writing. The reason we publish is to give citizens the information they need to govern themselves in a democracy.

FactCheck.org

We are a nonpartisan, nonprofit "consumer advocate" for voters that aims to reduce the level of deception and confusion in U.S. politics. We monitor the factual accuracy of what is said by major U.S. political players in the form of TV ads, debates, speeches, interviews and news releases. Our goal is to apply the best practices of both journalism and scholarship, and to increase public knowledge and understanding.

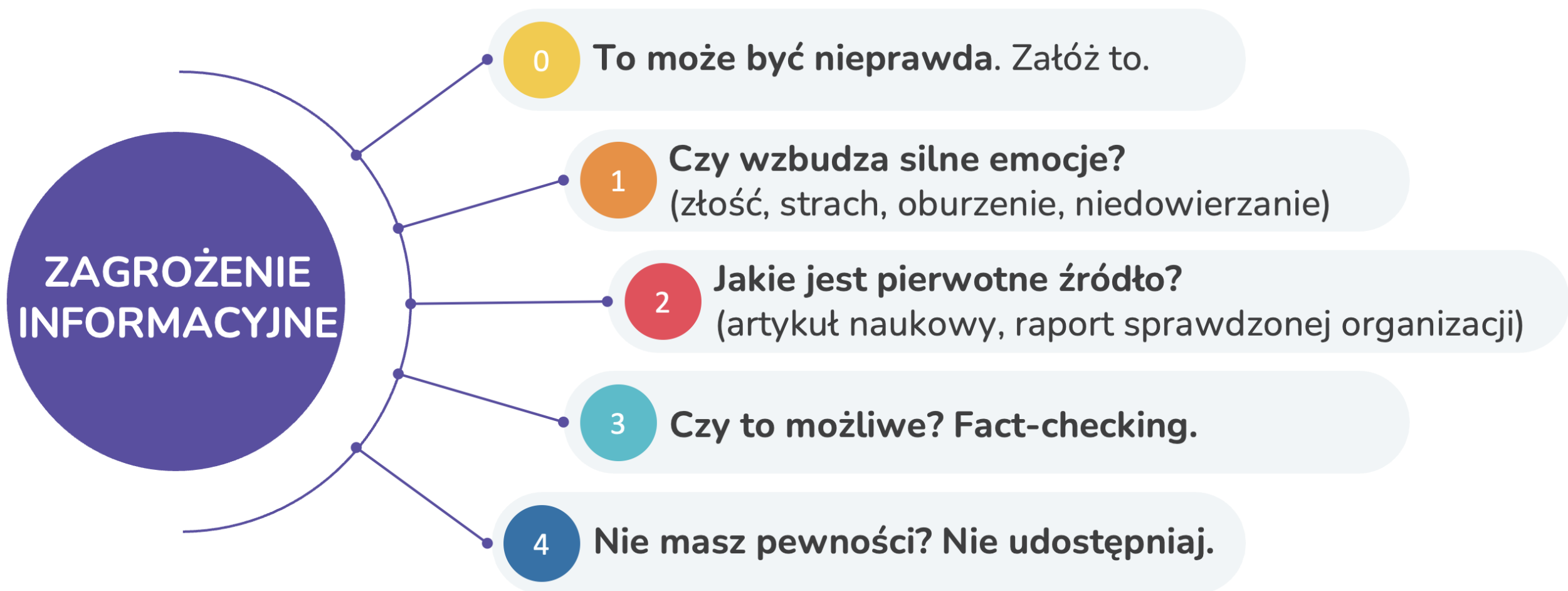
Washington Post Fact Checker

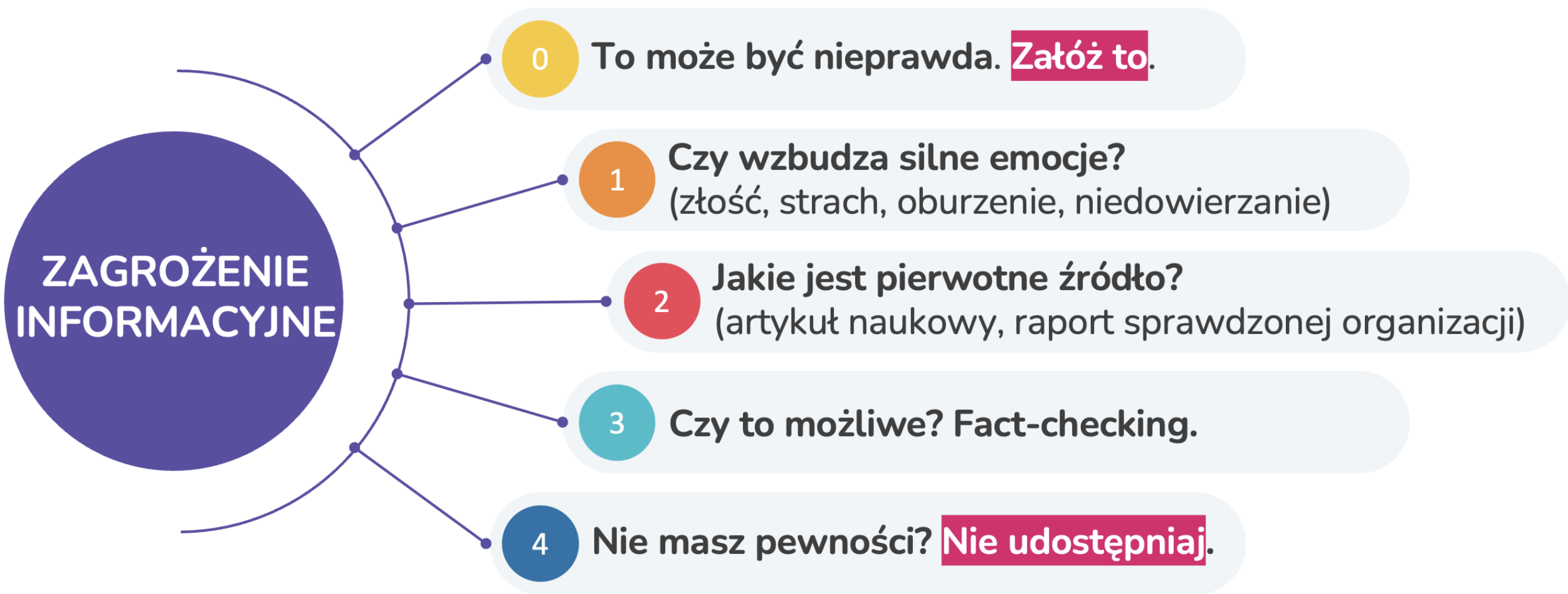
In an award-winning journalism career spanning nearly four decades, Glenn Kessler has covered foreign policy, economic policy, the White House, Congress, politics, airline safety and Wall Street. He was The Washington Post's chief State Department reporter for nine years, traveling around the world with three different Secretaries of State. Before that, he covered tax and budget policy for The Washington Post and also served as the newspaper's national business editor. He has been editor and chief writer of The Fact Checker since 2011.

Snopes



Reakcja na informację: 5 kroków





Między wsparciem a zagrożeniem: ciemna strona AI

JOANNA MYTNIK, CENTRUM NOWOCZESNEJ EDUKACJI POLITECHNIKI GDAŃSKIEJ