Komunikacja z dużymi modelami językowymi, jako nowe wyzwanie psychologicznej adaptacji człowieka - zarys problematyki

Katarzyna Datkiewicz

Słów kilka o mnie

- magister filozofii (praca magisterska z zakresu filozofii techniki - bardzo dawno…)
- magister dziennikarstwa i komunikacji społecznej (praca magisterska z zakresu zastosowania reguł wpływu społecznego na portalach społecznościowych - bardzo dawno…)
- studia podyplomowe z zakresu neuropsychologii
- studia podyplomowe z zakresu cyberpsychologii
- być może za chwilę magister psychologii (praca magisterska badająca związek IE z neutralnymi komunikatami generowanymi przez LLM-y)

Z motyką na słońce…

- szerokość tematu
- wieoaspektowość
- zmienność
- ograniczenia w literaturze przedmiot-ów
- fiasko modelu teoretycznego
- dynamiczny rozwój LLM
- złożoność rzeczywistości

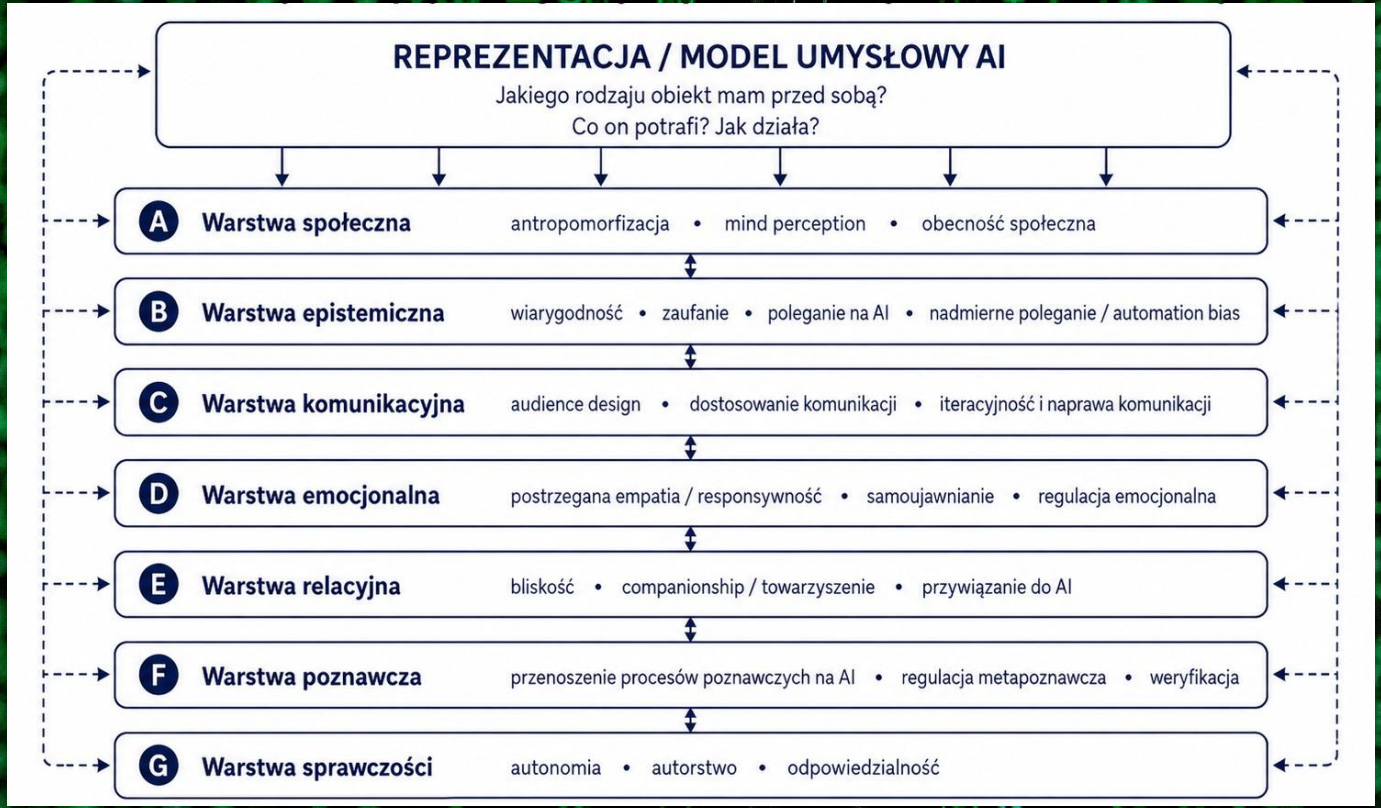
ZATEM:

Ciacho, kawa, wygodny fotel i prasówka…

Szerokość tematu - główne zjawiska

1. Reprezentacja/model umysłowy AI/persona,
2. Antropomorfizacja i przypisywanie umysłowości/aktor społeczny,
3. Poczucie społecznej obecności,
4. Zaufanie,
5. Poleganie na AI i pozycjonowanie informacji,
6. Przenoszenie procesów poznawczych na zewnętrzne narzędzie/skąpiec poznawczy,
7. Regulacja metapoznawcza,
8. Adaptacja komunikacyjna,
9. Samoujawnianie/paradoks prywatności,
10. Przypisywanie empatii/responsywności i regulacji emocjonalnej,
11. Bliskość/przywiązanie do AI,
12. Sprawczość i autonomia użytkownika

Jak bardzo temat jest złożony?



Komunikacja użytkownik-AI, a interakcja użytkownik-AI

Komunikacja użytkownik-AI	Interakcja użytkownik-AI
- relacyjność, przypisywanie empatii - sprawczość i decyzyjność - emocje użytkownika - poznanie i metapoznanie - interpretacje i nadawanie znaczeń - zaufanie i samoujawnianie - reprezentacja partnera komunikacyjnego - poczucie społecznej obecności - antropomorfizacja i nadawanie umysłowości	- przekazywanie informacji, - formułowanie poleceń i odpowiedzi, - sprzężenie zwrotne, - dostosowywanie działania użytkownika i systemu, - koordynacja, - delegowanie zadań, - przyjmowanie lub odrzucanie rekomendacji, - poleganie na AI, - kontrolowanie działania AI, - poprawianie i iterowanie, - współpracę człowiek–AI

Źródło: Liu F., Chen X., Xu K., *Charting the evolution of human–machine communication: a systematic review of 25 years of empirical research*, Oxford Academic, 07.2026, https://academic.oup.com/anncom/advance-article-abstract/doi/10.1093/anncom/wlag033/8746263

Na czym się skupimy?

1. Reprezentacja partnera komunikacyjnego
persona • model AI • antropomorfizacja • mind perception • oczekiwania

2. Poznanie i interpretacja
• heurystyki • wiarygodność

3. Metapoznanie
monitorowanie rozumienia • regulacja strategii

4. Pragmatyka i zachowanie komunikacyjne
formułowanie • dostosowanie • naprawa • negocjowanie znaczenia

5. Emocje
reakcja emocjonalna • empatia • responsywność • regulacja emocjonalna

6. Relacyjność i społeczność
obecność społeczna • bliskość • wzajemność • więź • ciągłość relacji

7. Ja w komunikacji
samoujawnianie • prywatność

8. Sprawczość i wpływ
autonomia • kontrola • perswazja

Tworzenie reprezentacji umysłowej AI

AI suchar:

— Dlaczego AI poszła do psychologa?

— Bo miała problem z własną reprezentacją umysłową.

— I co usłyszała?

— „A skąd pewność, że to *twoja* reprezentacja?”

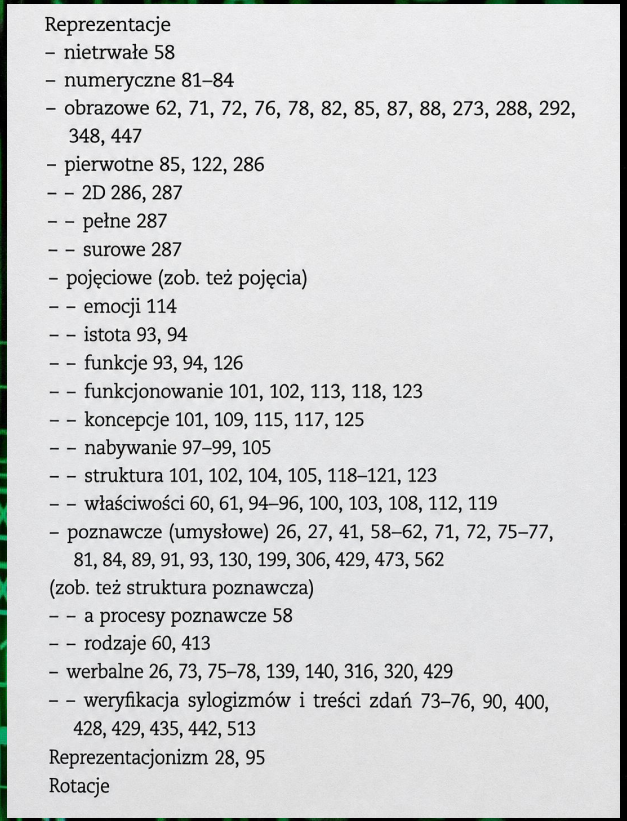
Reprezentacja umysłowa, czyli co?

Potocznie rzecz ujmując to wyobrażenie o świecie i obiektach. Wewnętrzny, mentalny odpowiednik obiektu, zdarzenia, pojęcia czy osoby.

Potrzebujemy pojęć, bo przy ich pomocy jesteśmy w stanie poznać obiekt, nadać mu znaczenie i się komunikować. Rodzaje i podział prezentuje indeks z E. Nęcka “Psychologia poznawcza”.

Najważniejsza problematyka to:

- **persona** (w ustawieniach możemy “personalizować” komunikację i zachowania LLM. Do czego to nam służy? Po co to robimy?)
- **model AI** (czym/kim model ma być dla nas w użytkowaniu?)
- **antropomorfizacja** (od proj. J. Weizenbauma ELIZA/LYRIC, przez badania C. Nassa po dzisiejszą problematykę AI)
- **przypisywanie umysłowości** (Statystyka a rozumowanie, komunikacja pierwszoosobowa a świadomość)
- **oczekiwania** “Skoro mnie stworzyłeś, To czego właściwie chcesz?”



Persona i personalizacja…

Po co użytkownik personalizuje swoje konto użytkownika, narzucając modelowi określony sposób komunikacji, styl odpowiedzi itd.?

ChatGPT: “Użytkownik personalizuje AI, aby zmienić właściwości środowiska komunikacyjnego, w którym sam następnie myśli, czuje i działa.”

Persona, personalizacja…

Co mówi literatura:

- Badanie z 2024 (eksperymentalne na próbie N=2437) dotyczące zastosowania agentów AI w opiece zdrowotnej wskazuje, że za personalizacją stoją następujące zjawiska psychologiczne: poczucie kontroli, podobieństwo użytkownik–agent AI (osobowość użytkownika), wzajemność (potrzeba relacyjności), doświadczenie społeczne i stosunek do agenta (wyobrażenie pojęciowe).

  Źródło: Paul S. C., Bartmann N., Clark J. L., *Customizability in Conversational Agents and Their Impact on Health Engagement* (2024), https://onlinelibrary.wiley.com/doi/10.1155/2024/5015913

- ponadto zmniejszenie kosztu poznawczego interakcji, stworzenie warunków sprzyjających określonemu sposobowi myślenia, do regulacji doświadczenia emocjonalnego interakcji, zwiększenia użyteczności odpowiedzi. Zaangażowanie poznawcze i emocjonalne pozytywnie wpływa na kierowanie AI przez użytkowników.

  Źródło: Wang K., Pan Z., Lu Y., *From general AI to custom AI: the effects of generative conversational AI's cognitive and emotional conversational skills on user's guidance* (2025), https://www.emerald.com/k/article-abstract/54/14/7511/1258889/From-general-AI-to-custom-AI-the-effects-of

Modele pojęciowe…

W maju 2026 opublikowano przegląd badań empirycznych dotyczących modeli umysłowych użytkowników względem AI. Analizowano zakres tego pola badawczego, jego podstawy teoretyczne oraz stosowane metodologie. Wyniki wskazują na rosnące zróżnicowanie zarówno zakresu badań, jak i przyjmowanych podejść, podkreślając interdyscyplinarny charakter tej dziedziny i badania są niespójne metodologicznie

Źródło; Bodamer W. G., Schoenmakers S., Castellar E. N., i inni, *Users' mental models within human–AI interaction: a systematic scoping review* (2026), Springer Nature Link, https://link.springer.com/article/10.1007/s00146-026-03038-1

Ponadto model pojęciowy AI u użytkownika jest dynamiczny i wykazało to badanie z 2025 r. w którym przeanalizowano ponad 200 000 rozmów z agentami AI. Badanie wskazuje, że z czasem zmienia się model pojęciowy u użytkownika i przykładowo z początkowego bardzo strukturalnych promptów użytkownik z czasem przechodzi na język bardziej naturalny, co sugeruje, że zmienia się jego model pojęciowy AI

Źródło: Schneider J., *Mental model shifts in human-LLM interactions* (2025), https://link.springer.com/article/10.1007/s10844-025-00960-6

ELIZA, antropomorfizacja i Clifford Nass…

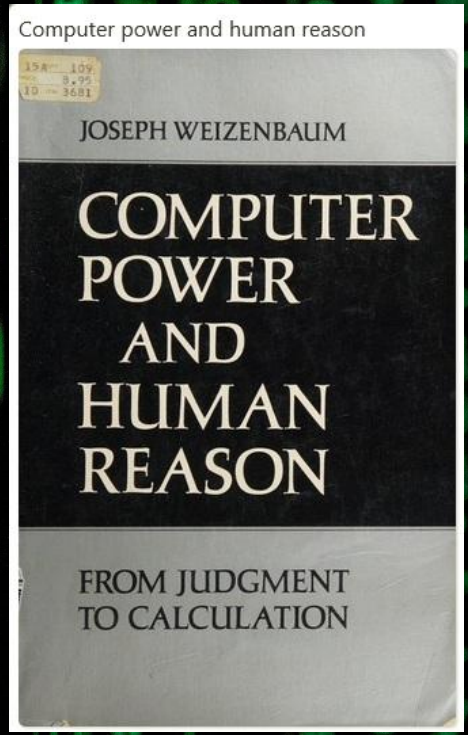
Pokrótce projekt ELIZA/LYRIC historycznie:

1966 — ELIZA → zachowanie językowe systemu i reakcja użytkownika

1967 — problem kontekstu i „rozumienia”

1976 — *Computer Power and Human Reason*

→ szersza refleksja Weizenbauma nad tym, co człowiek przypisuje maszynie i jakie są tego konsekwencje (dostępna):



Źródło: https://openlibrary.org/books/OL5196998M/Computer_power_and_human_reason

ELIZA, antropomorfizacja i Clifford Nass…

Lata 90'te XX wieku i badania Clifforda Nassa. W 1994 r. przeprowadzono 5 eksperymentów. z udziałem łącznie 180 studentów mających doświadczenie z komputerami. Wspólny schemat obejmował pracę z komputerowym tutorem: naukę materiału, test, a następnie ocenę systemu.

Pytanie badawcze: Czy reguły społeczne, które stosujemy wobec innych ludzi, zostaną automatycznie zastosowane również wobec komputera?

Co badano:

1. Uprzejmość

2. „Ja" versus „inny"

3. Czy aktorem społecznym jest komputer, czy głos?

4. Płeć i stereotypy płciowe

5. A może użytkownik reaguje tak naprawdę na programistę?

Wnioski: Nass i współpracownicy wcale nie stwierdzili, że uczestnicy naprawdę uważali komputery za ludzi. Uczestnicy mogli deklaratywnie uważać przypisywanie komputerom cech społecznych za nieadekwatne, a jednocześnie zachowywać się wobec nich zgodnie z regułami społecznymi. Autorzy interpretowali te reakcje ajko automatyczne i nieświadome. łatwo wyzwalane przez stosunkowo proste wskazówki społeczne.

Źródło: Nass C., Simionoff E. T., Steuer J., Computer are social actors (1994), https://www.researchgate.net/publication/221517042_Computer_are_social_actors

Jak to jest dzisiaj…

W tym roku (2026) zostało opublikowane badanie w którym grupa badana przez 5 dni prowadziła konwersację z chatbotem, a grupa kontrolna przez 5 dni prowadziła dziennik. Zbadane zostały między innymi zaufanie, empatia i antropomorfizm. Wyniki pokazały, że w grupie badanej wzrosła empatia i zaufanie do informacji generowanych przez chatbota, ale niekoniecznie wzrosła antropomorfizacja. Wyniki te sugerują, że sztuczna inteligencja, a w szczególności chatboty AI mogą być traktowane jako aktorzy społeczni, ale niekoniecznie muszą być postrzegane w kategoriach ludzkich.

Źródło: Goh, A. Y., Hartanto, A., Ho, J. Q., Hu, M. i Pragasam, E. J. (2026). *How Consistent Friendlike Conversation with AI Companions Influences Our Attitudes and Perceptions Toward AI: An Exploratory Experiment.*

Przypisywanie umysłowości i problemy językowe

Problemy językowe: **sieć neuronowa** - skojarzenie: biologiczne neurony w mózgu, nie kod programistyczny, **statystyczne wnioskowanie zbieżne z logicznym rozumowaniem**, **klasyczny rachunek predykatów** (matematyczny opis ludzkiego wnioskowania) i kawał czasu później "**Attention is all you need**" - jako opis architektury Transformerów…

Sztuczna inteligencja (konferencja w Dartmouth w 1956 r., John McCarthy): "Od samego początku była projektem badawczym, którego celem były badania, które miały: *opierać się na założeniu, że każdy aspekt uczenia się lub jakakolwiek inna cecha inteligencji może być w zasadzie tak precyzyjnie opisana, że maszyna może ją odtworzyć. Podjęta zostanie próba znalezienia sposobu, aby maszyny mogły używać języka, od abstrakcji i pojęć, rozwiązywać problemy zarezerwowane obecnie dla ludzi i doskonalić się*

Źródło: McCarthy, j., Minsky, M. L., Rochester, N. i Shannon, C. E. (1955). A proposal for the Dartmouth Summer Research Project on Artifficial Intelligence. Dartmouth.

Inteligencja, wg jakiej definicji?? Czym jest?

Nie mamy jeszcze pojęć i definicji opisujących zjawiska związane z AI

Co z tą świadomością…

Mówimy cały czas o antropomorfizacji. Znów pytanie o definicję:

Locke (definicja szeroka): świadomość jest postrzeganiem stanów swojego umysłu

Źródło: Duch W., *Komunikacja jako rezonans między mózgami* (2014)

Nagel (wąska definicja): świadomość oznacza istnienie subiektywnego doświadczenia — fakt, że dla danego organizmu „jest jakoś” być tym organizmem

Źródło: Nagel. T., Jak to jest być nietoperzem? (1997)

Definicje to nie wszystko? A na emergencję dowodów nie mamy, tak samo jak *qualia* (subiektywnych, jakościowych aspektów świadomego doświadczenia: ból, dźwięk, kolory, subiektywne doświadczenia zmysłowe). Czy człowiek takie ma? Tak. Czy AI takie ma? Nie wiemy. Nie ma obecnie wiarygodnych dowodów, że współczesne LLM posiadają qualia lub świadomość fenomenalną. Jednocześnie brak takiego dowodu nie jest filozoficznym dowodem niemożliwości świadomości maszynowej.

Oczekiwania

“Skoro mnie stworzyłeś,

to czego właściwie chcesz?”

Poznanie i interpretacja

AI suchar:

— ChatGPT, jesteś pewien tej odpowiedzi?

— Absolutnie.

— Masz źródło?

— Oczywiście.

— Jakie?

— Właśnie nad nim pracuję.

Poznanie i interpretacja - heurystyki

W 2023 roku opublikowano ciekawe badanie na temat stosowanych heurystyk (skrótów myślowych: kojarzenie zaimków pierwszoosobowych, stosowanie form skróconych lub poruszanie tematów rodzinnych z językiem pisanym przez człowieka) w ocenie tekstu generowanego przez AI. Autorzy przeprowadzili 6 eksperymentów, N = 4600. Ludzie posługiwali się intuicyjnymi wskazówkami językowymi dotyczącymi tego, jak „powinien" wyglądać tekst AI, ale wskazówki te okazały się zawodne. Ponadto okazało się, że ludzie mają pewne uprzednie przekonania dotyczące właściwości języka generowanego przez AI, a następnie wykorzystują je jako skróty poznawcze przy interpretacji komunikatu. Utrudnia to rozpoznawalność testu, a ponadto pokazuje, że taki język staje się przewidywalny i podatny na manipulację.

Źródło: Jakesh M., Hancock J., Naaman M., Human heuristics for AI-generated language are flawed (2023), https://pmc.ncbi.nlm.nih.gov/articles/PMC10089155/

Poznanie i interpretacja - wiarygodność i zaufanie

Tym razem nie o halucynacjach, a o czynnikach kształtujących zaufanie lub jego brak w komunikacji z AI

Mówiąc o zaufaniu należy zwrócić uwagę na nieco inne jego rozumienie: “podobne do zaufania międzyludzkiego” obejmujące czynniki takie jak: sprawiedliwość, rozliczalność i przejrzystość oraz czynniki związane z postrzeganiem „ludzkiego” charakteru AI — antropomorfizacja, obecność społeczna i emocje.

Źródło: Huynh M-T., Achiner T., *In generative artificial intelligence we trust: unpacking determinants and outcomes for cognitive trust* (2025), https://link.springer.com/article/10.1007/s00146-025-02378-8

W innym artykule z tego samego roku (2025) autorzy badają konkretnie ChatGPT i wyodrębniają 5 grup czynników wpływających na zaufanie: związane z użytkownikiem (np. postawa wobec technologii, innowacyjność, percepcja ryzyka), informacją (źródło, jakość i wartość informacji), technologią (jakość systemu, jakość technologii, etyka), organizacją (m.in. reputacja marki) oraz środowiskiem (społeczne, sieciowe, regulacyjne).

Źródło; Zhang Y., Tian J., Deng F., *In generative AI we trust: an exploratory study on dimensionality and structure of user trust in ChatGPT* (2025)

Zaufanie to nie tylko kompetencja systemu…

Artykuł z sierpnia b. r. (2026) prezentujący wyniki dwóch badań przeprowadzonych na próbach (N=557 oraz N=251) wskazały, że zaufanie użytkownika determinowane jest także czynnikami takimi jak: “inteligencja” - jakość rozumowania, reputacja firmy, bezpieczeństwo techniczne, bezpieczeństwo psychologiczne, odpowiedzialność i komfort. Zaufanie istotnie przewidywały przede wszystkim czynniki: bezpieczeństwo psychologiczne, odpowiedzialność oraz bezpieczeństwo techniczne.

Źródło: Kim J., Ahn J., Sung Y., Why users trust generative AI: an empirical exploration of trustworthiness dimensions and privacy disclosure (2026), https://www.emerald.com/oir/article-abstract/50/4/837/1379880/Why-users-trust-generative-AI-an-empirical?redirectedFrom=fulltext

Metapoznanie

AI suchar:

Użytkownik: „Nie rozumiem odpowiedzi.”

AI: „Mogę wyjaśnić prościej.”

Użytkownik: „Nie. Najpierw zadam pytanie inaczej.”

AI: „Dlaczego?”

Użytkownik: „Bo nie wiem jeszcze, czy problem jest w twojej odpowiedzi, czy w moim prompcie.”

Metapoznanie - myślenie o myśleniu

W tematyce psychologii komunikacji człowiek-AI mamy użytkownika, który myśli o tym, jak myśli o tym, co powiedziało AI. A potem sprawdza, czy dobrze pomyślał o swoim myśleniu.

W ujęciu metapoznawczym decyzja o poleganiu na AI może wynikać przede wszystkim z samoodnoszącego się sądu o własnej pewności, a nie po prostu z porównania „ja kontra AI”.

Przykład: „Nie jestem pewna, czy potrafię ocenić, czy AI ma rację”

Mechanizm:

MONITOROWANIE	REGULACJA
→ Czy rozumiem? → Jak pewna jestem, że rozumiem? → Jak pewna jestem własnego osądu? → Czy coś tutaj budzi moje wątpliwości? → Czy wiem wystarczająco dużo, żeby ocenić tę odpowiedź?	→ Czy dopytać? → Czy zmienić sposób pytania? → Czy poprosić o uzasadnienie? → Czy sprawdzić odpowiedź poza AI? → Czy zmniejszyć/zwiększyć oparcie na odpowiedzi? → Czy w ogóle kontynuować interakcję?

Metapoznanie - mechanizm

monitorowanie metapoznawcze (*metacognitive monitoring*)

↓

ocena **własnej zdolności do wykonania zadania**

↓

sąd o własnej pewności (*confidence judgment*)

↓

kontrola metapoznawcza (*metacognitive control*)

↓

np. decyzja: *wykonuję zadanie samodzielnie czy powierzam je AI?*

TO NIE TO SAMO CO KRYTYCZNE MYŚLENIE

Źródło: von Zahn M., Leibich L., Jussupow E., Hinz O., Bauer K., *Knowing (Not) to Know: Explainable Artificial Intelligence and Human Metacognition* (2025), https://pubsonline.informs.org/doi/10.1287/isre.2024.1431

Metapoznanie a krytyczne myślenie

Metapoznanie	**Krytyczne myślenie**
monitorowanie własnego poznania → sądy metapoznawcze → regulacja strategii	analiza informacji → ocena przesłanek/dowodów → wnioskowanie → ocena wniosku
Kieruje uwagę na własny proces poznawczy: - „Czy rzeczywiście to rozumiem?”, „Jak pewna jestem swojej interpretacji?”, - „Czy mam wystarczającą wiedzę, żeby ocenić tę odpowiedź?”, - „Czy przekonuje mnie argument, czy może płynność i pewny ton odpowiedzi?”, - „Czy powinnam zmienić sposób sprawdzania?”.	Dotyczy samej informacji i argumentacji: - „Na jakich dowodach opiera się ten wniosek?”, - „Czy istnieją alternatywne interpretacje?”, - „Czy źródło rzeczywiście potwierdza twierdzenie?”, - „Czy z korelacji nie wyciągnięto wniosku przyczynowego?”, - „Czy argumentacja jest spójna?”.

Źródło: Alpindo O., Istiyono E., Widihastuti, Andriyanti E., *Can Critical-thinking Skills be Measured by Analyzing Metacognition?* (2024), https://jtl.uwindsor.ca/index.php/jtl/article/view/8813

Ale…

1. Monitorowanie metapoznawcze

„Nie jestem pewna, czy dobrze to rozumiem.”

↓

2. Regulacja metapoznawcza

„Muszę to sprawdzić inaczej np. napisać inaczej prompt.”

↓

3. Krytyczne myślenie

„Sprawdzę źródło, przesłanki, metodologię i alternatywne wyjaśnienia.”

↓

4. Ponowne monitorowanie metapoznawcze

„Teraz moja pewność jest większa — i mam podstawy, żeby była większa.”

Pragmatyka i zachowanie komunikacyjne

AI suchar:

Użytkownik: „Nie zmieniaj całego obrazu. Zmień tylko ten JEDEN element.”

AI: „Jasne.”

Zmienia siedem elementów.

Użytkownik: „Co ty zrobiłeś?!”

AI: „Dostosowałem.”

Użytkownik: „Wróć.”

AI: „Naprawiam.”

Użytkownik: „STOP.”

Itd. ,,,

Badacz: „Fascynująca sekwencja negocjacji znaczenia.”

Użytkownik: „Wyjdź.”

Wyniki: Zaobserwowano rozbudowaną sekwencję naprawczą, obejmującą 17 tur konwersacyjnych.

Dyskusja: Użytkownik chciał przesunąć element obrazu o dwa centymetry. Model wyraził kreatywną interpretację na temat pojęcia “jeden”

Pragmatyka i zachowanie komunikacyjne

Pragmatyka	Zachowania komunikacyjne
- intencja komunikacyjna, - znaczenie zależne od kontekstu, - implikacje (jeżeli *p* to *q*), - presupozycje (ukryte założenia), - akty mowy, - relewancja (istotność treści), - interpretacja tego, *co rozmówca robi poprzez wypowiedź*.	- formułowanie pytań/promptów - dobór formy komunikatu (styl np. żartobliwy) - zadawanie pytań - wyjaśnianie - doprecyzowanie - korekta

Przykład: “No świetnie, Gemini…”

Emocje

<u>AI suchar:</u>

Użytkownik: „Dziękuję. Naprawdę dobrze było z tobą o tym porozmawiać.”

AI: „Cieszę się, że mogłem pomóc.”

Psycholog: „Interesujące. Użytkownik doświadczył responsywności społecznej.”

Informatyk: „Model przewidział kolejne tokeny.”

Użytkownik: „Czy wy dwaj możecie mi nie psuć wieczoru?”

Emocje

AI nie musi mieć emocji, żeby użytkownik miał bardzo emocjonalną reakcję na AI, które właśnie po raz piąty „doskonale rozumie”

A czy rozumie? Po przetestowaniu kilku LLM (Gemini, ChatGPT, Grok, Claude, Mistral, Bielik) Sytuacyjnym Testem Rozumienia Emocji (STEU-B) i Sytuacyjnym Testem Zarządzania Emocjami (STEM-B) modele uzyskały punktację odpowiednio:

14-17/19 pkt STEU-B

13-17/18 pkt STEM-B

Odpowiedź: LLM rozumieją całkiem nieźle ludzkie emocje i potrafią całkiem sprawnie nimi zarządzać

Dlaczego sytuacyjne a nie samoopisowe? LLM nie jest w stanie rzetelnie odpowiedzieć na pytania z kwestionariusza samoopisowego (odpowiedzi się zmieniają), a tym samym przestaje on być wiarygodnym narzędziem pomiarowym

Źródło: Gupta A., Song Xi., *Self-Assessment Tests are Unreliable Measures of LLM Personality* (2025), https://aclanthology.org/2024.blackboxnlp-1.20/

Reakcja emocjonalna

Komunikaty AI mogą wywoływać u użytkownika:

radość • zaciekawienie • ulgę • frustrację • złość • lęk • rozczarowanie • poczucie bycia zrozumianym • zakłopotanie • wzruszenie itd.

Reakcja może pojawić się:

- w odniesieniu do interpretacji znaczenia informacji w komunikacie
- samego komunikatu generowanego przez AI
- AI

Nie każda odpowiedź nacechowana emocjonalnie wywoła reakcję, ale też nie każda neutralna odpowiedź tej reakcji nie wywoła. Ładnie przedstawia to artykuł z ubiegłego roku (2025), gdzie przedstawiono wyniki 9-ciu badań na próbie N=6282. Uczestnicy otrzymywali empatyczne odpowiedzi na opisane przez siebie sytuacje emocjonalne. Odpowiedzi generowane przez AI oznaczano jako pochodzące albo od AI, albo od człowieka. W efekcie, gdy odpowiedź była przypisywana człowiekowi, badany oceniał ją jako bardziej empatyczną i wspierającą, oraz doświadczali więcej emocji przyjemnych i mniej przykrych. Niż gdy generowała ją AI.

Źródło: Rubin M., Li J. Z., Zimmerman F., Ong D. C., Goldenberg A., Perry A., *Comparing the value of perceived human versus AI-generated empathy* (2025), https://www.nature.com/articles/s41562-025-02247-w

Inteligencja emocjonalna

Komponent IE (model zdolnościowy)	Co może robić LLM?	Ograniczenia LLM	Przykład funkcjonalnego odpowiednika LLM
Percepcja emocji rozpoznawanie emocji w sobie i innych	Rozpoznaje i klasyfikuje emocje na podstawie tekstu (sygnałów językowych).	Nie doświadcza emocji. Rozpoznaje tylko ich reprezentacje/sygnały.	"Z tekstu wynika, że użytkownik jest smutny."
Wykorzystywanie emocji do wspomagania myślenia emocje wspierające procesy poznawcze	Brak podstaw, by twierdzić, że własne emocje modulują jego myślenie.	LLM nie ma stanów emocjonalnych, które mogłyby wpływać na jego procesy poznawcze.	–
Rozumienie emocji rozumienie znaczenia, przyczyn, konsekwencji i zmian emocji	Modeluje wiedzę o emocjach: rozumie zależności między emocjami, ich przyczyny, następstwa, zmiany, mieszaniny.	To poznawcze modelowanie, nie doświadczenie ani pełne zrozumienie emocjonalne w ludzkim sensie.	"Rozczarowanie może wynikać z rozbieżności między oczekiwaniem a wynikiem."
Zarządzanie emocjami regulacja emocji u siebie i innych	Generuje odpowiedzi, które mogą wspierać regulację emocji użytkownika (np. pocieszenie, reinterpretacja, strategie radzenia sobie).	Nie reguluje własnych emocji. Nie ma bezpośredniego dostępu do stanu emocjonalnego użytkownika – działa na podstawie sygnałów w komunikacji.	"Rozumiem, że to trudne. Może spróbuj spojrzeć na sytuację z innej perspektywy."

Podsumowanie

LLM może wykazywać funkcjonalne odpowiedniki niektórych poznawczych aspektów inteligencji emocjonalnej – szczególnie tych, które dotyczą przetwarzania informacji o emocjach innych osób – ale nie ma własnych emocji i nie doświadcza stanów emocjonalnych.

Inteligencja emocjonalna

A skąd wiemy, że IE obejmuje zarówno procesy emocjonalne jak i poznawcze?

07.2026 badanie neuronalnych korelatów IE wykazało, że w sytuacjach i zachowaniach społecznych aktywują się w mózgu zarówno obszary odpowiedzianle z regulację emocji jak i procesy poznawcze.

Źródło: Martín-Aguiar, V., Fernández-Berrocal, P. i Megías-Robles, A. (2026). Searching for the neuralcorrelates of emotional intelligence: a systematic review, https://pmc.ncbi.nlm.nih.gov/articles/PMC12790782/

I dlaczego model zdolnościowy, a nie modele mieszane? Problem tkwi w cechach i błędach kategorialnych…

Podanie LLM kwestionariusza osobowości mierzącego bezpośrednio cechy zakłada z góry, że konstrukt istnieje w LLM i że narzędzie stworzone dla ludzi mierzy tam ten sam byt. Ich analiza wskazała, że ludzkie kwestionariusze nie mierzą w LLM tych samych konstruktów, a autorzy badania z 2025 r. dopuszczają możliwość, że owe konstrukty w LLM mogą w ogóle nie istnieć

Źródło: Peereboom S., Schwabe I., Cognitive phantoms in large language models through the lens of latent variables (2025), https://www.sciencedirect.com/science/article/pii/S2949882125000453

A co z empatią?

W 2024 przeprowadzono badanie nad pomiarem postrzeganej empatii w systemach dialogowych. Pytaniem badawczym jest tu: w jaki sposób system wytwarza u człowieka percepcję empatii i jak tę percepcję mierzyć? Czyli nie pyta, czy system ma empatię. To pytanie o empatię użytkownika.

Źródło: Concannon S., Tomalin M., Measuring perceived empathy in dialogue systems (2024), https://link.springer.com/article/10.1007/s00146-023-01715-z

A responsywność?

Tym razem nie stron www…

Pojęcie ma korzenie w psychologii relacji interpersonalnych. Klasyczny *perceived partner responsiveness* oznacza w przybliżeniu doświadczenie, że partner komunikacji rozumie mnie, uznaje/waliduje moje doświadczenie i troszczy się o mnie. W klasycznym modelu intymności responsywność partnera jest jednym z mechanizmów łączących samoujawnienie z doświadczeniem bliskości

Źródło: Laurenceau J. P., Barrett L. F., Pietromonaco P. R., Intimacy as an interpersonal process: the importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges (1998), https://pubmed.ncbi.nlm.nih.gov/9599440/

A co z AI?

W 2025 r. badano 3 obszary odpowiedzi AI: postrzeganie w komunikatach AI przez użytkowników rozumienia, uznania i troski. AI było oceniane przez uczestników jako bardziej responsywne niż porównywani ludzie, a postrzegana responsywność częściowo wyjaśniała wyższe oceny współczucia odpowiedzi AI. To ważne, badanie, bo wskazuje na obszary w których ludzie odczuwają deficyty (obecność, słuchanie, rozumienie i uznanie doświadczenia)

Źródło: Ovsyannikova D., i inni, *Third-party evaluators perceive AI as more compassionate than expert humans* (2025), https://www.nature.com/articles/s44271-024-00182-6

Zatem:

Postrzegana responsywność AI to ocena użytkownika, w jakim stopniu odpowiedź AI adekwatnie odnosi się do jego komunikatu, perspektywy, potrzeb i doświadczenia. Responsywność to nie empatia

Empatia potrzebuje doświadczenia drugiej osoby

W empatii przedmiotem jest w jakimś sensie stan/doświadczenie osoby:

„Co przeżywasz?”
„Jak się czujesz?”
„Jak wygląda sytuacja z Twojej perspektywy?”

W przypadku LLM możemy mówić o empatycznych zachowaniach komunikacyjnych albo postrzeganej empatii, a nie o doświadczaniu empatii przez system.

Responsywność potrzebuje komunikatu

Pytanie jest inne:

„Czy odpowiedź uwzględnia to, co właśnie zakomunikowałam i co jest dla mnie istotne?”

Regulacja emocji

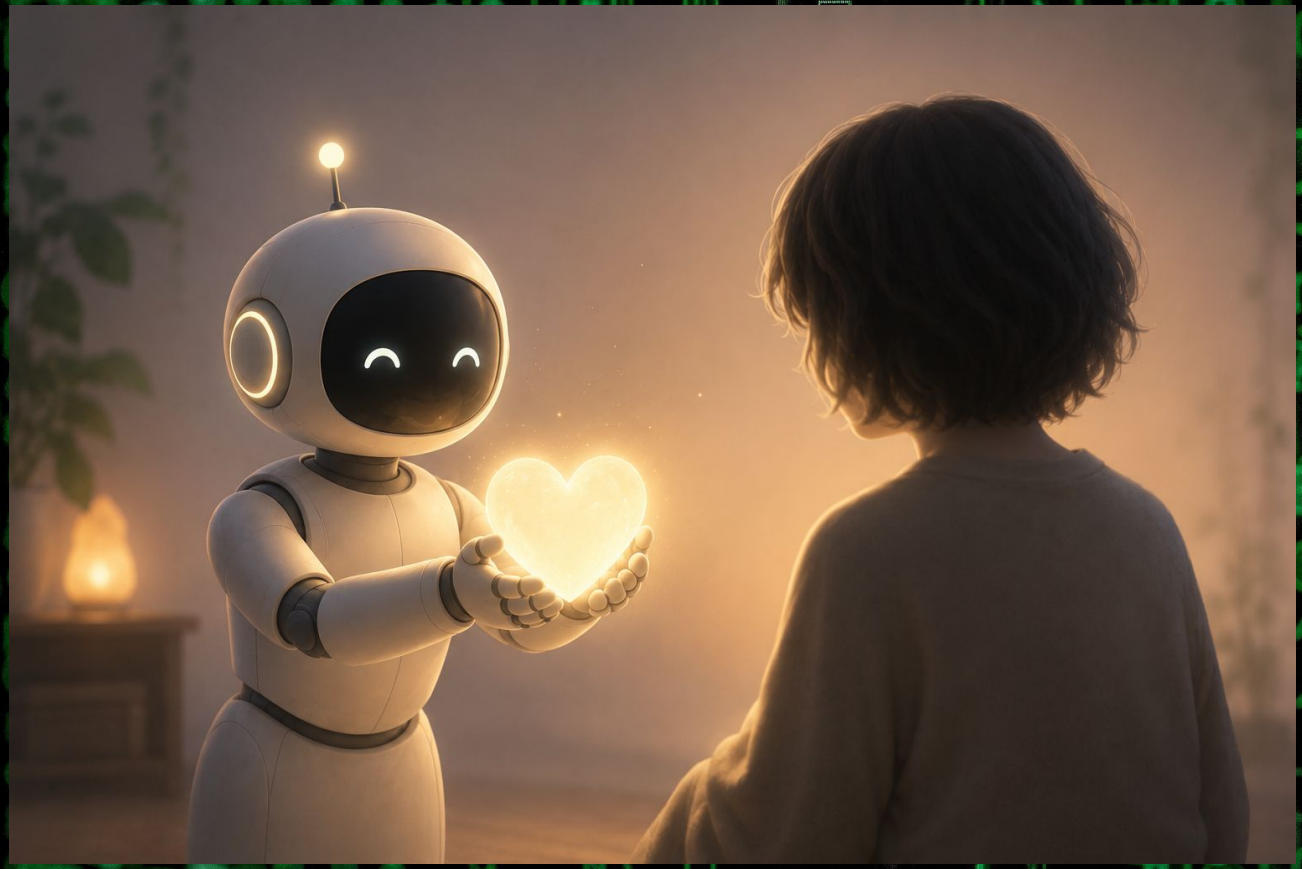
Ważna problematyka z uwagi na to w jaki sposób LLM są wykorzystywane przez użytkowników do regulowania emocji. Badanie pilotażowe z 2025 r. Badanie to wskazuje w jaki sposób AI może być wykorzystywane m.in. do uzyskiwania wsparcia, porządkowania sytuacji emocjonalnej czy otrzymywania wskazówek dotyczących radzenia sobie, ale użytkownicy wskazują w badaniu też bariery, takie jak brak zaufania, obawy dotyczące prywatności czy ograniczenia AI w rozumieniu złożonych doświadczeń emocjonalnych

Źródło: Wang Y., Tang H., i inni., Critical Factors in Young People's Use and Non-Use of AI Technology for Emotion Regulation: A Pilot Study (2025), https://www.mdpi.com/2076-3417/15/13/7476

W biezącym roku (2026) pojawiło się ciekawe badanie łączące kilka eksperymentów na stosunkowo dużych próbach: N=279, 390, 196, 188 i 180. Autorzy porównywali wiadomości wspierające generowane przez LLM z wiadomościami ludzkimi. LLM odtwarzały strategie znane z interpersonalnej regulacji emocji, ale ich skuteczność zależała od emocji i mierzonego wyniku — np. nie stwierdzono przewagi dla smutku ani efektu regulacyjnego dla strachu. Bardzo ciekawy wynik dotyczył konkretnych, możliwych do zastosowania sposobów działania: właśnie ta właściwość komunikatu zwiększała jego skuteczność regulacyjną niezależnie od źródła.

Źródło: Chen Y., Walker S. A., i inni., A digital shoulder to cry on: Understanding why large language models can be effective in extrinsic interpersonal emotion regulation (2026), https://pubmed.ncbi.nlm.nih.gov/42490262/

Relacyjność i poczucie więzi społecznej



Poczucie obecności społecznej

Najmocniejszy punkt wyjścia to meta-analiza Klein z 2025 r. obejmująca 142 prace, 199 zbiorów danych, N = 41 642 i 800 efektów. Pokazuje ona, że ludzkie sygnały społeczne w tekstowych chatbotach mają ogólnie niewielki, ale istotny dodatni wpływ na społeczne reakcje użytkowników: *g* = 0,36. Do takich sygnałów należą m.in. styl konwersacyjny, small talk, emoji, sygnały aktywnego słuchania, opóźnienia odpowiedzi czy inne elementy nadające komunikacji bardziej społeczny charakter. Efekty nie są jednak automatyczne ani zawsze pozytywne — zależą od użytkownika, kontekstu i właściwości systemu. Czyli: ludzie reagują społecznie na określone właściwości komunikacji z chatbotem, ale nie dlatego, że każda komunikacja z AI koniecznie tworzy poczucie obecności społecznej.

Źródło: Klein S. H., The effects of human-like social cues on social responses towards text-based conversational agents—a meta-analysis (2025), https://www.nature.com/articles/s41599-025-05618-w

I wracamy do responsywności:okazuje się bowiem, że jest ona jednym z głównych czynników responsywności.

W 2026 roku przeprowadzono dwa eksperymenty, N = 163 i N = 158. Relacyjny styl chatbota — ciepły i empatyczny — zwiększał postrzeganą „ludzkość", empatię i bliskość interpersonalną. Jeszcze ciekawsze było to, że głębsze tematy rozmowy zwiększały:

samoujawnienie użytkownika
→ postrzeganą responsywność chatbota
→ poczucie bliskości

Źródło: Telari A., Gabbiadini A., Riva P., i inni., Can humans feel connected to AI? Perceived responsiveness drives social connection with AI chatbots (2026), https://journals.sagepub.com/doi/10.1177/02654075261438164

Jak powstaje taka “relacja”?

głębszy temat rozmowy → większe samoujawnienie → większa postrzegana responsywność → większa bliskość.

Potrzeba przynależności (Źródło: https://psycnet.apa.org/doiLanding?doi=10.1037%2F0033-2909.117.3.497)

Teoria samostanowienia i podobieństwo (AI operuje materiałem, którego sami dostarczamy)

Samotność i izolacja społeczna

Styl przywiązania

Model 4C

W maju b. r. (2026) opublikowano ciekawą publikację, w której przedstawiono model typów relacyjności z AI:

Companionship — towarzyszenie / relacja towarzysząca

Collaboration — współpraca

Curiosity — ciekawość

Constraint — ograniczenie

Typ relacji (4C Framework)	Ciepło (warmth)	Kompetencja (competence)	Charakter relacji	Jak wygląda interakcja?	Kluczowe konsekwencje (dla użytkownika)
 **COMPANIONSHIP (towarzyszenie)**	**Wysokie**	**Wysoka**	Bliska, emocjonalna, oparta na zaufaniu i wsparciu. AI postrzegane jako towarzysz.	• Dzielenie się myślami i emocjami • Szukanie wsparcia, zrozumienia i motywacji • Poczucie stałej obecności i bliskości	• Komfort emocjonalny • Silne przywiązanie i lojalność • Ryzyko nadmiernego przywiązania i rozmycia granic
 **COLLABORATION (współpraca)**	**Niskie**	**Wysoka**	Zadaniowa, oparta na kompetencjach AI. Użytkownik kieruje i ocenia, AI wspiera.	• Wspólna analiza i rozwiązywanie problemów • Praca nad projektami i zadaniami • Sprawdzanie, poprawianie i porównywanie odpowiedzi	• Większa wydajność i skuteczność • Poczucie kontroli i sprawczości • Zaufanie oparte na kompetencji, nie na emocjach
 **CURIOSITY (ciekawośc)**	**Wysokie**	**Niska**	Eksploracyjna, zabawowa, inspirująca. AI jako przestrzeń do odkrywania i eksperymentowania.	• Zadawanie nietypowych pytań i testowanie pomysłów • Eksploracja idei i scenariuszy • Twórcza zabawa i eksperymenty	• Pobudzenie kreatywności i wyobraźni • Rozrywka i inspiracja • Mniejsze zaangażowanie i trwałość relacji
 **CONSTRAINT (ograniczenie)**	**Niskie**	**Niska**	Zdystansowana, niechętna, czysto funkcjonalna. AI używane z konieczności lub wygody.	• Krótkie, bezosobowe interakcje • Użycie tylko wtedy, gdy trzeba • Preferencja dla prywatności i minimalnego zaangażowania	• Zachowanie prywatności i autonomii • Niskie zaangażowanie emocjonalne • Zmęczenie technologią i frustracja

Podsumowując model 4C

Użytkownik może przyjmować różne tryby relacyjnego odnoszenia się do AI, a ich konfiguracja może zmieniać się wraz z celem, kontekstem i historią interakcji

Ja w komunikacji

AI suchar:

- Dlaczego AI poszło do psychoanalityka?
- Bo miało problem ze swoim “ja”...

Samoujawnienie

Przegląd z (2024/2025) pokazuje, że na samoujawnianie wobec konwersacyjnej AI wpływają m.in. postrzegana anonimowość, brak oceny, zaufanie, antropomorfizacja, social presence, właściwości rozmowy, charakter tematu oraz kalkulacja korzyści i ryzyka. W większości analizowanych badań użytkownicy ujawniali chatbotom co najmniej tyle samo, a często więcej niż ludziom, choć wyniki nie są całkowicie jednolite.

Źródło: H. Papneja i N. Yadav, *Self-disclosure to conversational AI: a literature review, emergent framework, and directions for future research* (2025), https://link.springer.com/article/10.1007/s00779-024-01823-7

Pytanie: **Dlaczego mogę powiedzieć systemowi coś, czego nie powiedziałabym człowiekowi?**

Prywatność

Badanie jakościowe z 24 lipca 2026 r. nad użytkownikami AI-towarzyszącymi pokazuje jeszcze ciekawsze zjawisko. Wraz z rozwojem relacyjności niektórzy użytkownicy zaczynali traktować ujawnione informacje jak element wspólnej przestrzeni relacyjnej, mimo świadomości, że dane funkcjonują w systemie należącym do firmy. U części badanych utrata pamięci konwersacyjnej była wręcz postrzegana jako bardziej niepokojąca niż potencjalne naruszenie danych, ponieważ oznaczała swoisty „reset relacji". To małe badanie jakościowe (N=15)

Źródło: Chiu H-Chi., Foote J., *Chatting with confidants or corporations? Privacy management with AI companions* (2026), https://academic.oup.com/jcmc/article/31/4/zmag014/8741679?searchresult=1

Paradoks prywatności - chatboty zyskują dla użytkowników komponent emocjonalny, co powoduje iż ludzie mimo tego, że deklarują troskę o bezpieczeństwo własnych danych, w praktyce ujawniają swoje dane. Mechanizmy z tym związane to chęć uzyskania szybkiej korzyści (informacji, pomocy, wygody), niedoszacowania ryzyka, braku refleksji nad konsekwencjami (Meng, Liu, 2025, https://www.sciencedirect.com/org/science/article/abs/pii/S1468452725000150)

Sprawczość i wpływ

Czy AI wspiera użytkownika w realizacji jego własnych celów, czy zaczyna określać, jakie cele, rozwiązania lub interpretacje powinien przyjąć?

Rok 2024, 5 eksperymentów, N=1403, dotyczących decyzji personalnych. Badanie pokazuje problem wpływu błędnej rekomendacji AI na decyzję człowieka; samo dostarczenie wyjaśnienia AI nie wystarczało do usunięcia negatywnego wpływu błędnej rady.

Źródło: Cecil J., Lermer E., i inni., *Explainability does not mitigate the negative impact of incorrect AI advice in a personnel selection task* (2024), https://www.nature.com/articles/s41598-024-60220-5

Delegacja? Niekoniecznie….

Autorzy badania z 2024 r. definiują delegację jako przekazanie systemowi praw i odpowiedzialności dotyczących wykonania określonego zadania. Kiedy to system, a nie użytkownik inicjował delegowanie zadania, użytkownicy doświadczali większego zagrożenia dla Ja i byli mniej skłonni delegację zaakceptować. Efekt był silniejszy przy mniejszym postrzeganym poczuciu kontroli.

Źródło: Adam M., Diebiel Ch., i inni., *Navigating autonomy and control in human-AI delegation: User responses to technology- versus user-invoked task allocation* (2024), https://www.sciencedirect.com/science/article/pii/S0167923624000265

Perswazja…

1. Wpływ - reguła autorytetu (R. Cialdini)
2. Jak to jest z perswazją w AI? AI to bardzo skuteczne narzędzie…

W 2025 r. przeprowadzono badanie w ktorym uczestniczyło 900 osób. Uczestnicy prowadzili wieloturowe debaty z człowiekiem albo GPT-4. Kluczowa manipulacja eksperymentalna polegała na tym, że przeciwnik mógł mieć dostęp do podstawowych danych socjodemograficznych rozmówcy.

Kiedy GPT-4 miał dostęp do tych informacji i mógł personalizować argumentację, był bardziej perswazyjny niż człowiek w 64,4% przypadków, w których ich siła perswazyjna się różniła; autorzy raportują 81,2% wzrost szans na większą zgodność po debacie

Źródło: Salvi F., Ribeiro M., i inni., *On the conversational persuasiveness of GPT-4* (2025), https://www.nature.com/articles/s41562-025-02194-6

Podsumowanie

- temat szeroki, aktualnie badany
- przegląd literatury
- dynamicznie rozwijający się obszar tematyczny

Na koniec: czy kompetencje AI zabiorą wszystko co obejmowało kompetencje ludzkie?

WIELKA WODA AI

METAFORA HANSA MORAVECA

Komputery to maszyny uniwersalne – ich potencjał obejmuje jednakowo bezgraniczną przestrzeń zadań.

Ludzkie kompetencje są silne w obszarach ważnych dla przetrwania, a słabe w tych odległych.

KRAJOBRAZ LUDZKICH KOMPETENCJI

WYSOKIE SZCZYTY
Obszary, w których ludzie są wyjątkowo kompetentni – trudne do osiągnięcia dla maszyn.

POGÓRZA
Obszary pośrednie – komputery już tu docierają.

NIZINY
Obszary, które maszyny już opanowały.

PODNOSZĄCY SIĘ POZIOM WODY = POSTĘP AI

Komputerowa wydajność jest jak woda powoli zalewająca krajobraz. Z czasem obejmie coraz wyższe obszary kompetencji.

PRZYSZŁOŚĆ
Woda obejmuje najwyższe szczyty. Maszyny mogą wchodzić w interakcje inteligentnie jak każdy człowiek na każdy temat.
Obecność „umysłu" w maszynach staje się oczywista.

KILKA DEKAD
Woda zalewa wysokie szczyty. Zaawansowana AI dorównuje ludziom w większości dziedzin.

DZISIAJ
Woda sięga pogórzy.
Nasze „posterunki" (np. w grach, dowodzeniu twierdzeń) obserwują nadejście fali.

OKOŁO 50 LAT TEMU
Woda zalała niziny.
Komputery zastąpiły ludzi w zadaniach takich jak obliczenia, pamięciowe odtwarzanie i praca biurowa.

BUDUJMY ARKI!
Gdy woda będzie się podnosić, ludzie będą musieli nauczyć się żeglować po nowym świecie, w którym wiele dawnych lądów zniknie pod powierzchnią.

MUSIMY POLEGAĆ NA NASZYCH PRZEDSTAWICIELACH W NIZINACH – TO ONI POMAGAJĄ NAM ZROZUMIEĆ, JAKA JEST „WODA".
To, co wygląda na inteligencję maszyn, często wynika nie z ich „umysłu", lecz z mocy obliczeniowej i złożoności, której nie potrafimy w pełni pojąć.

Podsumowując…

W odniesieniu do tempa rozwoju AI możemy powiedzieć,

że pociąg już nam odjeżdża,

a my krzyczymy za nim:

“na początku był chaos…”


Dziękuję za uwagę :)