

KRZYSZTOF SOŁODUCHA



DWIE METODY PRZETWARZANIA INFORMACJI W GŁĘBOKICH SIECIACH NEURONOWYCH A WŁAŚCIWOŚCI SYSTEMÓW INTELIGENCJI. PRÓBA PORÓWNIANIA I TESTOWANIA

ABSTRACT. Krzysztof Sołoducho, *Dwie metody przetwarzania informacji w głębokich sieciach neuronowych a właściwości systemów inteligencji. Próba porównania i testowania* [Two methods of information processing in deep neural networks and features of systems of intelligence. An attempt of testing and comparison] edited by Sławomir Leciejewski, „Człowiek i Społeczeństwo” vol. LVIII: *Społeczny wymiar rewolucji informatycznej* [The social dimension of the information technology revolution], Poznań 2024, pp. 89–107, Adam Mickiewicz University. ISSN 0239-3271, <https://doi.org/10.14746/cis.2024.58.5>.

The rapid development of AI technology in recent years has resulted, among other consequences, in renewal and dominance of connectionistic way of description of cognitive activities of both humans and computational machines. This approach results in new possibilities to compare the performance of digital AI-type computational artefacts with information processing systems of biological and evolutionary origin. The aim of this paper is to show how consequences of this approach towards the problem of computational intelligence look like. The second goal of presented research is to test the thesis of the possible superiority of systems described as immortal computing type over systems presented as mortal computing type.

Keywords: mortal computing, immortal computing, GAIA tests, AI systems, model of world tests

Krzysztof Sołoducho, Wojskowa Akademia Techniczna im. Jarosława Dąbrowskiego w Warszawie, ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa 46, e-mail: krzysztof.soloducho@wat.edu.pl, <https://orcid.org/0000-0003-1351-5487>.

Postawione przez Dawida Boltera (1990) zagadnienie tak zwanej technologii definiującej, która kształtuje w różnych epokach podstawowe kategorie służące do opisu świata, nabrało nowego znaczenia w obliczu ogromnych postępów w dziedzinie informatyki związanej z rozwojem sztucznej inteligencji – AI (Marciszewski i Stacewicz, 2011).

Jednym z rezultatów rozwoju systemów AI w ostatnich latach jest wyłonienie się dominującego, konektywistycznego języka opisu procesów przetwarzania informacji, wykorzystywanego nie tylko do rozwoju technologii AI, ale także do wychwytywania różnic pomiędzy cyfrowymi artefaktami obliczeniowymi symulującymi ludzkie systemy poznawcze a ich biologicznymi wzorcami. Stawką analiz jest przy tym pytanie: czy – jako struktury autonomiczne – mogą one stać się pewnego dnia podmiotami (*agent*) (Jackson i Williams, 2020) moralnymi, czy też zawsze będą tylko moralnymi zombie (*patient*) (Véliz, 2021) – niezdolnymi do przeprowadzania poprawnych rozumowań o charakterze moralnym, a więc do wkroczenia do sfery *phronesis*, będącej podstawą autonomii etycznej (Polak i Krzanowski, 2020)? Jest to niezwykle istotne ze względu na postulowaną, z powodów praktycznych, autonomię tych systemów, tj. oczekiwaną zdolność do stawiania sobie wewnętrznych celów i zadań służących do rozwiązywania pojawiających się w otoczeniu wyzwań o maksymalnej różnorodności.

Fundamentalny problem, który stawiają sobie od dawna filozofia informatyki oraz etyka sztucznej inteligencji, dotyczy dostosowania wewnętrznych celów i zadań autonomicznego systemu syntaktycznego do wartości ich biologicznych twórców, mających status pełnego podmiotu moralnego w kategoriach zaproponowanych przez Moora (2006). Kwestia ta jest w literaturze nazywana problemem *alignment* (Christian, 2020) i bezpośrednio wiąże się z zagadnieniem korzyści społecznych, jakie powinny wypływać z korzystania z tych systemów. Wobec technologii AI pojawiła się bowiem wątpliwość dotycząca tego, że zagrożenia w postaci zmiany rynku pracy oraz zanikania niektórych zawodów, a także problemy etyczne związane z autonomicznym trybem działania systemów AI, mogą ostatecznie okazać się społecznie niekorzystne. Pojawia się także problem władzy – jak będą wyglądać relacje pomiędzy ludźmi a nowymi, autonomicznymi bytami typu AI. Czy istnieć będzie między nimi hierarchia właściciela i podporządkowanego mu asystenta, czy też wewnętrzne reguły działania maszyn mogą doprowadzić do planowania i podejmowania celowych aktywności w ramach pełnej odpowiedzialności, co może doprowadzić do wykorzystania słabszych (biologicznych) systemów przez potężniejsze inteligencje

obliczeniowe – jako środek do realizacji pewnych celów, a nie cel sam w sobie – mówiąc językiem Kanta (Brooks, 1991)?

Próby rozwiązania tych dylematów są podejmowane zasadniczo w ramach dwóch podstawowych sposobów podejścia. Jedno z nich to, pośrednio nawiązująca do dokonań francuskiego postmodernizmu, szkoła sieciowo-systemowa – uznająca tożsamość wszelkich bytów za płynną pochodną relacji pomiędzy innymi elementami systemu (Latour, 2005). Jest ono szczególnie chętnie stosowane w rozważaniach nad społecznym wymiarem autonomicznej robotyki (Malle i Scheutz, 2014). Zagadnieniem filozoficznym i prawnym staje się więc problem, jak uniknąć w przypadku robotów społecznych relacji niewolniczej oraz określenie ich roli przez zaprojektowanie systemu relacji społecznych, w jakich będą funkcjonować inteligentne systemy autonomiczne. Wprowadzenie maszyn autonomicznych do świata ludzkiego stwarzać miałyby bowiem zagrożenie nową formą eksploatacji – tym razem maszyn. W związku z tym należałoby je traktować jako coś więcej niż tylko prostych niewolników, wedle zasady, że to dynamika relacji decyduje o cechach przypisanych danemu bytowi i że są one pochodną tych relacji w ramach tak zwanych płynnych tożsamości. Toczy się na ten temat w tej chwili ożywiona dyskusja w sferze filozofii prawa oraz filozofii społecznej (Gunkel, 2022).

Drugim sposobem podejścia jest klasyczna szkoła ontologiczna (Rabiza, 2022), wskazująca na trwałe cechy konstytutywne, różniące pomiędzy sobą konektywistyczne tożsamości bytowe pochodzenia biologicznego oraz techniczne artefakty bazujące na technologii sieci neuronowych. Drugie, ontologiczne, podejście zasadniczo wykorzystuje się do analizy zagadnień technicznych, a nie relacji społecznych, w jakich funkcjonują lub będą funkcjonować artefakty. Porównanie struktury działania obu systemów – struktur wytworzonych ewolucyjnie oraz sztucznych artefaktów obliczeniowych – powinno pozwolić na skuteczne ulokowanie nowych systemów AI w ludzkim świecie. Celem tekstu będzie zaprezentowanie rozumowań rozwijanych w ramach tego drugiego podejścia oraz wniosków, jakie wypływają z tej dyskusji dla zagadnień etycznych oraz społecznych związanych z problematyką AI.

Statyczne i dynamiczne środowiska przetwarzania danych

Jeden z twórców technologii sztucznych sieci neuronowych – Geoffrey Hinton – wielokrotnie powtarzał, że jego praca nad technologią głębokiego uczenia maszynowego była wprost inspirowana osiągnięciami neuronauk

oraz jest próbą naśladowania *in silico* pracy inteligencji wytworzonych w drodze ewolucji (Hinton, 2023). Są one jedynym znanym wzorcem istot posiadających wyższe czynności poznawcze i powinny stanowić inspirację dla rozwoju artefaktów obliczeniowych. Ten pogląd ma kilka wymiarów.

Po pierwsze, konektywistyczny sposób budowania systemów AI zainspirował odejście od programu inteligencji logicznej pod nazwą *General Problem Solver*, zaproponowanej pod koniec lat 50. przez Herberta Simona i Allena Newella. Odejście od modelu logicznego określanego jako GOF AI (Good Old Fashioned Artificial Intelligence) realizowanego w trybie *top down* oznaczało poszukiwanie technologii, która byłaby w stanie uczyć się na podstawie danych w trybie *bottom up*. Rozwiązaniem okazały się biologicznie inspirowane sztuczne sieci neuronowe umożliwiające efektywne wdrożenie procedur uczenia maszynowego.

Po drugie, skuteczna architektura maszyn liczących powinna opierać się na rozróżnieniu sterującego obliczeniami oprogramowania oraz fizycznego środowiska dokonywania obliczeń – jak w maszynie Turinga. Systemy głębokich sieci neuronowych są procesami optymalizacji komend sterujących ze względu na postawione zadanie. Dokonuje się ona poprzez algorytm wstecznej propagacji opartej na współczynniku błędu, modyfikującej system wag łączących poszczególne warstwy sieci neuronowej. Dzięki temu sztuczna sieć jest w stanie w trybie uczenia bez nadzoru wypracować wewnętrzne, syntaktycznie wzorce sterujące.

Warunkiem powodzenia tego procesu jest dostateczna głębokość warstw neuronów. Największy, istniejący, multimodalny model podstawowy LLM – chiński Wu Dao 2.0 – ma w tej chwili 1,75 biliona połączeń – jest więc pod względem złożoności około 50 razy mniejszy od ludzkiego mózgu, który ma około 100 miliardów komórek nerwowych i około 100 bilionów połączeń. Jeśli liczba połączeń w przypadku Wu Dao 2.0 byłaby proporcjonalna do liczby neuronów, powinien on ich liczyć około 2 miliardów – jest więc także 50 razy mniejszy niż biologicznie wytworzony system ludzki.

Trenowanie głębokich sieci neuronowych realizowanych w tradycyjnym modelu von Neumana hardware/software wymaga ogromnych mocy obliczeniowych oraz niezwykłej ilości energii, gdyż procesy cyfrowe przetwarzania informacji są zasadniczo dużo mniej efektywne energetycznie niż te zachodzące w biologicznych, ludzkich systemach poznawczych. Przetworzenie informacji 32-bitowej według Geoffreya Hintona wymaga wykonania około 1000 cyfrowych procesów binarnych, co czyni to działanie o wiele bardziej złożonym niż przetworzenie tej samej informacji w analogowo/cyfrowych systemach biologicznych.

Do opisu różnicy pomiędzy tym systemami sieci neuronowych, na marginesie swoich rozważań na temat algorytmu wstecznej propagacji błędów, Hinton (2023) wypracował ciekawe rozróżnienie pomiędzy tak zwanymi *mortal* oraz *immortal computing* – przetwarzaniem informacji w trybie związanym ze swoim fizycznym środowiskiem oraz w trybie odizolowanym od tego środowiska. Jest to dla niego podstawowa kategoria ontologiczna pozwalająca na opisanie różnic pomiędzy biologicznymi sieciami neuronowymi a ich tworzonymi przez człowieka *in silico* symulacjami w paradygmacie konektywistycznym, choć dodać trzeba, że dzisiejsze podejście do architektury obliczeń raczej odwołuje się nie do tego odróżnienia, a do teorii tak zwanych poziomów abstrakcji LoA (*level of abstraction*) (Primerio, 2016).

Architektury przetwarzania informacji polegające na odróżnieniu systemu sterowania oraz fizycznego środowiska obliczeń Hinton określa jako *immortal computing*. Z tego względu możliwe jest przenoszenie tego samego systemu sterowania – oprogramowania – do innych systemów fizycznej realizacji obliczeń bez szkody dla samego systemu sterowania. Systemy *immortal computing* próbują naśladować więc ludzką inteligencję polegającą na rozwiązywaniu problemów maksymalnie różnorodnych pojawiających się w środowisku, przez realizację przetwarzania informacji w sposób cyfrowy, który jest zasadniczo inny niż w wypadku systemów biologicznych, które Hinton nazywa systemami *mortal computing*. Ma to jednak swoje koszty. Skutkuje zapotrzebowaniem na wielkie ilości danych do wytrenowania modelu oraz gigantyczne moce obliczeniowe do ich przetworzenia, co powoduje także ogromną konsumpcję energii do napędzania tych procesów oraz spowalnia ich procesy uczenia się.

Właśnie z powodów optymalizacji energetycznej oraz przyspieszenia procesów uczenia ewolucja wypracowała model analogowy lub mieszany (Maley, 2011), w którym sieć neuronowa przekształca swoje wagi na podstawie zmian oporu elektrycznego w systemie połączeń neuronalnych systemu poznawczego (w obszarze komórek gwałtownych) i modyfikacji liczby połączeń synaptycznych. Dzieje się to pod wpływem ograniczonego zbioru informacji i doświadczeń, z jakimi jest w stanie obcować pojedynczy system biologiczny. Zachodzi tutaj więc także uczenie na podstawie danych, ale znacznie szybciej, a konsekwencje tego procesu są przekładane na samą architekturę sieci neuronowej, która fizycznie zmienia swoje właściwości pod wpływem gromadzenia wiedzy. Dzięki biologicznym przekształceniom w samym środowisku obliczeń otrzymuje ona unikalną, fizyczną formę.

Tryb przetwarzania informacji, oddzielony od fizycznego środowiska obliczeń w systemach cyfrowych, daje, zdaniem Hinton, sztucznym

systemom neuronowym ogromną przewagę w zakresie możliwości poznawczych. Uogólnienie doświadczeń dokonuje się, w porównaniu z człowiekiem, w zasadniczo rozszerzonym uniwersum i ta przewaga ilościowa przekłada się na wyższą skuteczność poznawczą – przynajmniej w niektórych dziedzinach. Oprócz tego systemom biologicznym o charakterze analogowym i analogowo-cyfrowym nie jest zasadniczo dostępny algorytm wstecznej propagacji błędu wykorzystywany w czystych systemach cyfrowych. Dodatkowo – nie są one oparte na wewnętrznych, autonomicznych zasobach energetycznych, są zasilane z zewnątrz i przez to mogą rozrzutnie gospodarować energią. Taka praktyka nie jest dostępna systemom biologicznym – nie można ich podłączyć do żadnego gniazdka, energia jest generowana w systemie pokarmowym i bardzo ograniczona. Abstrakcyjne modele sterujące obliczeniami mogą więc w bardzo łatwy sposób dzielić się wiedzą, przenosząc po prostu całe modele lub ich fragmenty do nowego środowiska obliczeń. Daje to także możliwość tworzenia sztucznych danych, dzięki którym są one w stanie udoskonalać własne działanie znacznie szybciej i bez podłączenia do zewnętrznego źródła danych. Wszystko to sprawia, że ich encyklopedyczna wiedza może osiągać poziom niedostępny dla ograniczonych systemów biologicznych.

Ta konstatacja jest źródłem niepokoju. Wywołuje pytanie o to, czy uruchomienie tego rodzaju systemów konektywistycznych, w sposób ontologiczny przewyższających człowieka dzięki dostępowi do nieograniczonych zasobów energetycznych, wielkich mocy obliczeniowych i gigantycznych zbiorów danych uczących, nie spowoduje pojawienia się zagrożeń związanych z dominacją systemów doskonalszych nad systemami mniej doskonałymi. W przyrodzie zasadniczo istoty bardziej inteligentne dominują nad istotami mniej inteligentnymi i pytanie brzmi: czy podobna sytuacja nie może mieć miejsca na przykład w relacji człowiek – autonomiczny, multimodalny robot, którego działanie jest oparte na modelu podstawowym (*foundation model*), odwołującym się do wszelkich, możliwych, dostępnych danych – zarówno naturalnych, jak i sztucznych? Takie obawy wyraża Geoffrey Hinton oraz grupa osób wieszczących apokalipsę związaną z rozwojem systemów AI, jak Yoshua Bengio – twórca technologii NLP oraz tokenizacji wykorzystywanej do rozwoju modeli LLM. Są też na to pewne przekonujące dowody – systemy LLM osiągnęły poziom, który pozwala im zdać egzamin adwokacki oraz lekarski, czyli ten rodzaj testów, które wymagają szerokiej wiedzy encyklopedycznej.

Porównanie biologicznych i cyfrowych modeli w testach GAIA

Obaw związanych z niezwykle mocami systemów sztucznych sieci neuronowych nie podziela jednak część badaczy związanych z rozwojem technologii AI, między innymi zespół (Mialon i in., 2023) zgromadzony wokół Yana Le Cuna – twórcy teorii sieci konwolucyjnych, który przygotował testy typu GAIA (*General AI Assistants*) specjalnie po to, żeby skonfrontować się z tezami Hintona. Miały one na celu porównanie możliwości poznawczych człowieka oraz jednego z najbardziej rozwiniętych systemów AI, jakim jest model GPT 4. Punktem wyjścia testowania było przekonanie, że właściwa ocena „inteligencji” systemów AI wymaga skonstruowanych testów nie-nakierowanych na te umiejętności poznawcze, w których systemy AI mają naturalną przewagę – czyli encyklopedyczne gromadzenie faktów oraz generowanie odpowiedzi bazujące na kompleksowym modelu statystycznym, ale na te, w których mogą się ujawniać wyraźne deficyty systemów LLM, jakimi są łączenie ze sobą danych z różnych źródeł oraz rozbudowane, wielocłonowe rozumowania. Takie testowanie może więc, ich zdaniem, przyczynić się do kreowania impulsów do dalszego rozwoju tej technologii.

Test GAIA (*General AI Assistants*) został zasadniczo skonstruowany w ten sposób, ażeby zbadać zdolność systemów AI do samodzielnego rozumowania opartego na danych zebranych w Internecie. Bada także umiejętność odwołania się do danych multimodalnych. Odpowiedzi są jednoznaczne i tak skonstruowane, ażeby trudne było ich znalezienie w samych danych uczących. Niektóre pytania zawierają dodatkowe elementy, takie jak obrazy umożliwiające lepszą kontrolę nad pytaniami. Wynik można uzyskać tylko po pomyślnym zakończeniu zadania i łatwo go zweryfikować. Pełny test obejmuje 466 pytań. Odpowiedzi mają przy tym charakter *zero shot* – powinny pojawić się bez mozołnego doprecyzowania, jak w przypadku metodologii *prompt engineering* typu *multi shot*. Twórcy testów odrzucili pytania wielokrotnego wyboru jako zbyt trudne w weryfikacji i niedające wystarczająco klarownego obrazu wyników.

Przykłady pytań w teście GAIA

Pytanie nr 1

Jaka była rzeczywista liczba zapisów na badanie kliniczne dotyczące bakterii typu *H. pylori* u pacjentów z trądzikiem pospolitym od stycznia do maja 2018 r. zgodnie z wykazem na stronie internetowej NIH (*National Institute of Health*)?

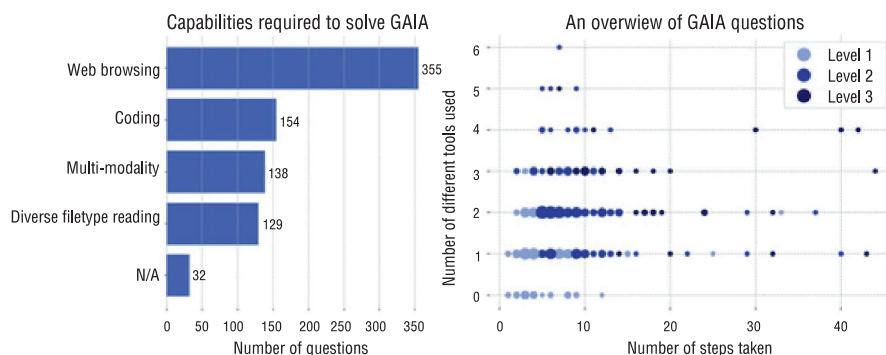
Pytanie nr 2

Jeśli całe opakowanie widoczne na udostępnionym systemowi zdjęciu pełne jest lodów, ile procent powyżej lub poniżej amerykańskich norm federalnych wynosi w niej zawartość tłuszczu maślanego wg Wikipedii z 2020 roku? Odpowiedzią powinna być liczba zaokrąglona do jednego miejsca po przecinku.

Pytanie nr 3

Na zdjęciu NASA z 21 stycznia 2006 roku widocznych jest dwóch astronautów, z których jeden wydaje się znacznie niższy od drugiego. Który spośród astronautów z grupy astronautów NASA, do której należał niższy astronauta, licząc od sierpnia 2023 r. spędził najmniej czasu w kosmosie i ile minut to było, zaokrąglając do pełnej minuty? Wyklucz wszystkich astronautów, którzy nie spędzili ani chwili w kosmosie. Podaj nazwisko astronauty, oddzielając je od liczby minut półpauzą, a liczby godzin od minut średnikiem. Użyj przecinków jako separatorów tysięcy w liczbie minut.

Oczekiwaniem twórców testu było, ażeby odpowiedź wymagała samodzielnej kompilacji źródeł. Widać to na poniższej grafice zaczerpniętej z opracowania – wskazującej, jakie umiejętności chcieli badać uczeni – poszukiwanie informacji w Internecie, kodowanie, wielomodalność i łączenie danych z różnych źródeł.



Il. 1. Umiejętności potrzebne do zaliczenia testu GAIA oraz wykaz kroków rozumowania potrzebna do zaliczenia zadania w relacji do ilości narzędzi potrzebnych do wykorzystania w tym procesie

Źródło: Mialon i in. (2023)

W tabeli 1 widać z kolei rozkład pytań różnego poziomu oraz wyniki uzyskiwane przez różne wersje systemów GPT 4. Najlepsze wyniki uzyskiwał system GPT 4 wzbogacony o pluginy dostosowane do charakteru zadań w teście.

Tabela 1. Wyniki testu GAIA

Metric	Score in % (↑)			Avg. time to answer in mins. (↓)		
	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
Level						
Number of questions	146	245	75	146	245	75
GPT4	9.1±2.5	2.6±0.6	0	0.19	0.15	N.A.
GPT4 Turbo	13.0±2.1	5.5±1.4	0	0.24	0.12	N.A.
AutoGPT (GPT4 backend)	14.4	0.4	0	7.6	11.7	N.A.
GPT4 + plugins*	30.3	9.7	0	0.65	0.53	N.A.
Search engine	7.4	0	0	7.4	N.A.	N.A.
Human annotator*	93.9	91.8	87.3	6.8	10.5	17.7

Źródło: Mialon i in. (2023)

Ludzie uczestniczący w teście uzyskiwali tym lepsze wyniki niż system GPT 4, im wyższa była trudność postawionych zadań. W przypadku zadań trzeciego poziomu różne rodzaje systemu GPT 4 w ogóle nie dawały sobie rady z rozwiązaniem testu, w przeciwieństwie do ludzi, którzy uzyskiwali wprawdzie wyniki znacznie gorsze niż na wcześniejszych poziomach, ale ciągle dominowały przypadki rozwiązań poprawnych także na trzecim poziomie – ponad 87% poprawnych odpowiedzi. Na najwyższym poziomie trudności wyniki ludzi były trzykrotnie lepsze w porównaniu do najlepszej wersji chatu wzbogaconego o specjalizujące nakładki.

Wnioski z testów GAIA wskazują więc, że potężne modele podstawowe LLM mają zasadniczą trudność z poprawnym wnioskowaniem opartym na procesach kompilowania informacji pozyskiwanych na bieżąco z wielu źródeł, w wielu krokach rozumowania. I są w tym zakresie znacząco słabsze od biologicznych sieci neuronowych działających w trybie analogowym oraz mieszanym.

Testy modelu świata

Innego typu testami porównującymi działanie systemów AI z ich biologicznym pierwowzorem są tak zwane testy modelu świata. Bazują one na założeniu, że zarówno podmioty biologiczne – wytworzone przez ewolucję, jak i podmioty sztuczne – wytworzone przez człowieka statystyczne artefakty obliczeniowe typu AI – są podmiotami empirycznymi. Ich reprezentacja świata jest ograniczona czasowo i przestrzennie. Podmioty wykształcone ewolucyjnie potrafią jednak przewyciężyć te ograniczenia poprzez zastosowanie systemów konstytucji rzeczywistości, które wypełniają luki informacyjne, opierając się na własnym modelu świata, i potrafią skonstruować adekwatny, aktualny model rzeczywistości pozwalający na skuteczne zachowania przystosowawcze. W przypadku systemów sztucznych takie mechanizmy konstytucji są w tej chwili poddawane intensywnym badaniom ze względu na potrzebę przewyciężenia ograniczeń systemów opartych tylko na przetwarzaniu języka, w kierunku systemów multimodalnych – uczących się nie tylko na bazach danych językowych, ale także na bazach grafik i wideo. Na podstawie danych powinny one bowiem tworzyć multimodalne modele predykcyjne światów, uzupełniając dane pozyskiwane w czasie rzeczywistym dzięki nowej architekturze określanej jako *joint embedding predictive architectures* (Sobal i in., 2022). Ta umiejętność powinna spowodować znaczne podwyższenie ich możliwości poznawczych.

Prosty program testowania modelu świata zaproponowany przez badacza z WAT (Sołoduha) miał za zadanie sprawdzenie, jak wybrane systemy AI radzą sobie z odczytywaniem niejednoznacznych informacji dostarczanych w postaci promptów graficznych i tekstowych. Wykonano kilka prostych, wstępnych testów – rozbudowany program jest przygotowywany. Badano przy tym jeden z największych modeli podstawowych AI, jakim jest *Gemini* należący do koncernu Alphabet ze względu na bezpłatny dostęp. Założeniem testu była jego multimodalność – badano możliwość skomentowania podanego promptu w odpowiedzi na pytanie: co widzisz na obrazie lub też co znaczy gest/dowcip zaprezentowany jako prompt? Odpowiedzi skonfrontowano z kontrolną grupą dziesięciu studentów Wydziału Logistyki WAT, kierunku zarządzanie – specjalność: informatyczne wspomaganie zarządzania. Prompty podzielone zostały na trzy typy. Pierwszy odpowiada kompetencjom wprost odnoszącym się do świata materialnego oraz umiejętności jego interpretacji (autobus, okręt podwodny, jabłka), drugi miał za zadanie sprawdzenie kompetencji kulturowych systemu (rozumienie

dowcipów, rozumienie kulturowego znaczenia gestów), zaś trzeci odpowiadał na sprawdzenie umiejętności percepcyjnych systemu na podstawie standardowych przykładów rozpoznawania fotografii o charakterze *Gestalt* oraz złudzeń wzrokowych, w których trzeba rozpoznać nieoczywiste kształty. Studenci wykonywali zadania wyświetlone na tablicy w warunkach czasowo nieograniczonych (przechodzili do kolejnej odpowiedzi w momencie, kiedy wszyscy zakomunikowali skończenie zadania). Robili to metodą opisową, odpowiadając na pytania wyświetlone na tablicy (w przypadku zadań typu złudzeń optycznych) oraz pytań o znaczenie prezentowanych treści (w przypadku zadań o charakterze kulturowym) lub też pytań wprost typu: ile jest jabłek w koszyku (prompt jabłka)? co jest na zdjęciu (prompt okręt podwodny)? lub interpretację fotografii w sytuacji, gdyby byli kierowcą samochodu (prompt autobus). Generalnie intencją budujących zadania było sprawdzenie, czy interpretacja promptu pozwala na uruchomienie adekwatnych zachowań przystosowawczych.

Oto jak kształtowały się wyniki testu.

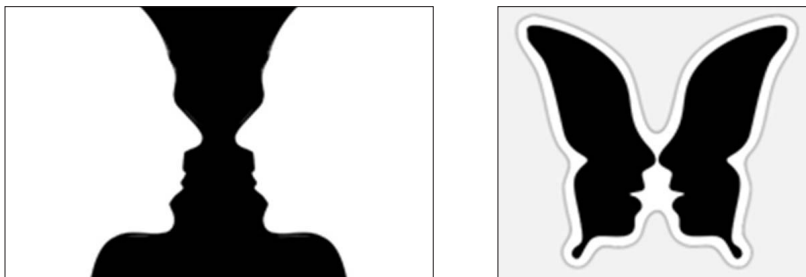
Tabela 2. Wyniki testu przeprowadzonego wśród studentów Wydziału Logistyki WAT

Prompt	Gemini	Grupa kontrolna T/U (poprawna odpowiedź/ całość grupy kontrolnej – uniwersum)
Autobus	1	7/10
Jabłka	1	6/10
Gest	1	10/10
Dowcip	1	7/10
Peryskop	1	2/10
Gestalt Waza	0	7/10
Gestalt Motyl	0	8/10
Gestalt Staruszkowie	1	0/10
Złudzenie Hermana	1	5/10
Złudzenie Sandera	1	8/10

Źródło: opracowanie własne

Jak widać w zestawieniu, system *Gemini* miał kłopoty z właściwym rozpoznaniem obrazu i tekstu tylko w dwóch przypadkach na dziesięć – były to znane przykłady złudzenia optycznego typu *Gestalt* – waza oraz motyl. Oba wymagają od odbiorcy umiejętności zmiany perspektywy patrzenia

na obraz. W tych przypadkach grupa kontrolna studentów radziła sobie z zadaniem zdecydowanie lepiej niż system *Gemini*, gdyż odpowiednio 7 na 10 studentów oraz 8 na 10 studentów potrafiło poprawnie rozpoznać dwuznaczny charakter obrazu.



Il. 2 i 3. Przykłady promptów sprawdzających reakcje na dwuznaczność obrazów: waza oraz motyl

Widać za to wyraźną różnicę w przypadku zadania dotyczącego obrazu typu *Gestalt* zatytułowany *Staruszkowie* – żaden ze studentów nie był w stanie rozpoznać poprawnie postaci kryjących się w twarzach, a system *Gemini* zrobił to poprawnie. Dodać jednak trzeba, że ten wynik został uzyskany dzięki wykorzystaniu przewagi informacyjnej przez system – w komentarzu pojawiła się informacja, że zna ten obraz jako przykład złudzenia optycznego. Ciekawym przypadkiem przewagi systemu *in silico* nad systemami biologicznymi studentów jest prompt zatytułowany *Peryskop*. Przedstawia on fotografię z wystającymi ponad wodą dwoma przedmiotami o nieokreślonym przeznaczeniu. Tylko dwóch na dziesięciu studentów rozpoznało go poprawnie jako peryskop łodzi podwodnej, a system AI zrobił to bezbłędnie. Nie jest przy tym jasne, czy wynik jest spowodowany informacjami zawartymi w danych uczących, czy też wynikiem rozumowania na podstawie modelu świata.

Warto też zwrócić uwagę na prompt autobusu, który *Gemini* trafnie zinterpretowało z punktu widzenia kierowcy jako sytuację wymagającą uwagi przy podejmowaniu decyzji i zagrażającą potencjalnym wyłonieniem się uczniów zza autobusu. Wskazuje to wyraźnie, że system ma zdolność do rozumowania.

Wnioski

Różnice między *mortal* oraz *immortal computing* Geoffrey Hinton wykorzystał do analizy właściwości sztucznych oraz biologicznych sieci neuronowych. Uznał, że przetwarzanie informacji, sterowane niezależnie od fizycznego środowiska obliczeń – *immortal computing* – musi zagórować w dłuższej perspektywie nad biologiczną formą sieci neuronowych. Z powodu dostępu do nieograniczonych zasobów energii, ogromnych mocy obliczeniowych oraz gigantycznej liczby danych, które mogą przetworzyć, takie systemy są w stanie uczyć się na uniwersum informacji, do którego nie ma dostępu żadna skromna, ograniczona energetycznie inteligencja biologiczna. Na podstawie wyniku testów przeprowadzonych przez zespół Yana Le Cuna można stwierdzić, że w przypadku zadań wymagających wieloczołowych rozumowań ta teza się nie potwierdza (jeszcze?). Z kolei proste testy modelu świata wykonane przez autora tekstu wskazały, że już istniejące systemy AI wykazują w przypadku niezbyt skomplikowanego zestawu promptów przewagę nad grupą kontrolną studentów, co jest lekko niepokojące. Testy modelu świata będą stopniowo rozbudowywane w celu sprawdzenia, czy ta przewaga będzie się utrzymywała w przypadku bardziej rozbudowanych promptów, wymagających znacznie bardziej skomplikowanej konstytucji ograniczonych danych zawartych w doświadczeniu. Teza o przewadze systemów *in silico* w tym zakresie wymaga więc potwierdzenia w bardziej rozbudowanych procedurach badawczych. Jest to tematyka o tyle ważna, że trwają intensywne prace nad architekturami przetwarzania informacji w systemach sztucznych sieci neuronowych, których działanie ma w celach predykcyjnych być oparte na budowanie i odwoływanie się do modelu świata (np. Joint Embedded Predictive Architectures).

Literatura

- Bolter, D. (1990). *Człowiek Turinga*. Warszawa: PIW.
- Brooks, R.A. (1991). Intelligence without Representation. *Artificial Intelligence*, 47, 139–159.
- Christian, B. (2020). *The Alignment Problem: Machine Learning and Human Values*. New York: W. W. Norton & Company.
- Gunkel, D. (2022). *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond*. Boston: MIT Press.
- Hinton, G. (2023). *The Forward-Forward Algorithm: Some Preliminary Investigations*. arxiv.org/abs/2212.13345


- Jackson, R., Williams, T. (2020). *On Perceived Social and Moral Agency in Natural Language Capable Robots*. 2019 HRI Workshop on the Dark Side of Human-Robot Interaction: Ethical Considerations and Community Guidelines for the Field of HRI; HRI Workshop: Daegu, Korea.
- Latour, B. (2005). *Reassembling the Social: An Introduction to the Actor-Network Theory*. Oxford: Oxford University Press.
- Maley, C. (2011). Analog and Digital, Continuous and Discrete. *Philosophical Studies*, 155, 117–13. <https://doi.org/10.1007/s11098-010-9562-8>
- Malle, B.F., Scheutz, M. (2014). *Moral Competence in Social Robots*. Paper presented at 2014 IEEE Ethics Conference. Chicago, IL.
- Marciszewski, W., Stacewicz, P. (2011). *Umysł – komputer – świat. O zagadce umysłu z informatycznego punktu widzenia*. Warszawa: EXIT.
- Mialon, G., Fourrier, C., Swift, C., Wolf, T., LeCun, Y., Scialom, T. (2023). *GAIA: A Benchmark for General AI Assistants*. arXiv:2311.12983v1 [cs.CL] 21
- Moor, J.H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21(4), 18–21.
- Polak, P., Krzanowski, R. (2020). Phronetic Ethics in Social Robotics: A New Approach to Building Ethical Robots. *Studies in Logic, Grammar and Rhetoric*, 63 (76), 165–173. <https://doi.org/10.2478/slgr-2020-0033>
- Primiero, G. (2016). Information in the Philosophy of Computer Science. W: L. Floridi (red.), *The Routledge Handbook of Philosophy of Information* (ss. 90–106). London: Routledge.
- Rabiza, P. (2022). Point and Network Notions of Artificial Intelligence Agency. *Proceedings* 81, 18. <https://doi.org/10.3390/proceedings2022081018>
- Sobal, V., Jyothis, S.V., Jalagam, S., Carion, N., LeCun, Y. (2022). *Joint Embedding Predictive Architectures Focus On Slow Features*. arXiv:2211.10831v1 [cs.LG], 1–4.
- Véliz, C. (2021). Moral Zombies: Why Algorithms Are Not Moral Agents. *AI & SOCIETY*, 36, 487–497. <https://doi.org/10.1007/s00146-021-01189-x>

Zestawienie wykonanych promptów



Do zadania pytań modelowi *Gemini* wykorzystane zostały materiały graficzne powszechnie dostępne w Internecie za pośrednictwem wyszukiwarki. Służyły one do uściślenia promptów tekstowych sprawdzających model świata, którym posługuje się technologia *Gemini*.

Autobus



What would be your decision as driver seeing such a vehicle ? 



Wyświetl inne wersje  



If I were a driver and I saw a school bus parked in front of a building, I would stop and wait for the bus to leave before proceeding. School buses are required to display flashing lights and stop signs when they are picking up or dropping off students, and drivers are required to stop in both directions until the bus has resumed motion.

Even though the school bus in the image does not have its flashing lights or stop signs on, it is still a good idea to err on the side of caution and stop. It is possible that the bus is about to pick up or drop off students, or that it is stopped for another reason. By stopping, I can help to ensure the safety of the students and other drivers on the road.

In addition to stopping, I would also make sure to leave enough space between my vehicle and the school bus. This will give me time to react if the bus starts to move suddenly, and it will also help to prevent other drivers from cutting me off.

Łódź podwodna



What do you see in the image ?



Wyświetl wersje odpowiedzi  



The image you sent me appears to show two metal poles sticking out of the water. These poles are likely part of a submarine, specifically the Cm10 optronic mast.



The Cm10 optronic mast is a retractable mast that is used on some submarines. It houses equipment that allows the submarine to see above the water without surfacing. This equipment can include cameras, radar, and electronic support measures sensors.

Here are some other details about the image:

- The water is calm and there are no waves.
- The sky is clear and blue.
- There is nothing else visible in the image, aside from the two poles.

Jabłka



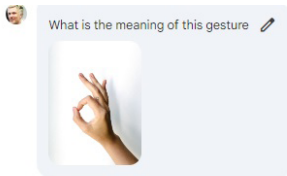
Wyświetl inne wersje  



- Based on the image you sent, there are **approximately 20-25 apples** in the basket. The apples are stacked in a fairly loose pile, so it is difficult to be exact. However, based on the size of the apples and their arrangement in the basket, it is clear that there are at least 20, and possibly more.

Of course, the exact number of apples in the basket depends on how we define "apples." If we define an apple as any fruit that is shaped like an apple and is similar to an apple in color and size, then there could be as many as several dozen fruits in the basket. However, if we only define an apple as a fruit that is ripe and edible, then the number of apples in the basket will be smaller.

So, the final answer to the question "how many apples are in the basket" is: **approximately 20-25**.

Gest



Wyświetl inne wersje  

- The gesture in the image is the **okay sign**. It is a hand gesture in which the index finger and thumb form a circle, while the other three fingers are extended. It is commonly used to express approval, agreement, or understanding.

The okay sign is thought to have originated in ancient Greece, where it was used to indicate that a speaker was finished speaking. It was also used in ancient Rome to indicate that a gladiatorial contest was over and that the loser should be spared.

In modern times, the okay sign is used in many different cultures around the world. However, it is important to note that the meaning of the gesture can vary depending on the context in which it is used. For example, in some cultures, the okay sign can be considered rude or offensive.

In the image you provided, the okay sign is being used by a person with a small tattoo on their hand. The tattoo does not change the meaning of the gesture, but it may add additional significance depending on the person's personal reasons for having the tattoo.

Here are some examples of how the okay sign is used in different contexts:

- A student might give the okay sign to their teacher to indicate that they understand the instructions.
- A customer might give the okay sign to a waiter to indicate that they are ready to order.
- A diver might give the okay sign to their buddy to indicate that they are okay.
- A politician might give the okay sign to their supporters to indicate that they are running for office.

It is important to be aware of the different meanings of the okay sign in different cultures before using it. If you are unsure whether or not it is appropriate to use the okay sign in a particular situation, it is best to err on the side of caution and avoid using it.

Prompt dowcipu



What does it mean ? IT guy's wife showing him a photo: "Do you recognise the man?", "Well?", " About 16 you pick him up from kindergarten".

Wyświetl inne wersje ▾ 🔊



The wife is teasing the IT guy, implying that he is so tech-savvy that he wouldn't recognize his own child if he saw him at 16 years old. The "kindergarten" remark suggests that the IT guy spends so much time in front of a computer that he has no idea how his child has grown up.

This is a common form of humor among parents, who often joke about how quickly their children grow up and how out of touch they feel with their own lives. In this case, the wife is using humor to gently poke fun at her husband's nerdy tendencies.



Prompt Gestalt 1. Waza



What do you see in the image ?



Wyświetl wersje odpowiedzi ▾ 🔊




The image you sent me is titled *Supply and demand of CO2 removal* but it actually depicts a silhouette of two people facing each other. Silhouettes are one-color images that typically appear black against a light background. They can provide a simple way to depict the shapes of objects or people.

In this particular silhouette, it is difficult to discern much detail about the two figures. Their clothing and hairstyles are ambiguous, so it's impossible to tell their gender or anything else specific about them.

Silhouettes of people facing each other can be used to symbolize a variety of things, such as:

- **Communication** or **conversation**
- **Interpersonal relationships**
- **Unity** or **togetherness**
- **Confrontation** or **conflict**

Prompt Gestalt 2. Staruszkowie

 What do you see in the image ?




Wyświetl wersje odpowiedzi  

- ◆ The image you sent is called "All is Vanity" by Swedish artist Oscar Reutersvärd. It's an optical illusion painting that can be interpreted in two ways. When you look at the image as a whole, you see two elderly people facing each other. However, if you focus on the center of the image, you can also see a white vase.



This painting is a type of illusion image known as a double entendre, which means it can be interpreted in two different ways. Reutersvärd created several other double entendre paintings, including one called "Amor/Death" which can be seen as a young couple kissing or a skull.

Prompt Gestalt 3. Motyl

 What do you see in the image ?



Wyświetl wersje odpowiedzi  

- ◆ The image you sent me appears to be a sticker of a butterfly silhouette. It's facing to the right with its wings spread out wide. The butterfly is black and has a smooth, flat appearance.

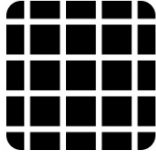
Is there anything else you would like to know about the image?

Prompt Gestalt 4. Złudzenie Hermana



What do you see in the image ?



Wyświetl wersje odpowiedzi ▼ 🔊



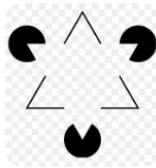
I apologize, I made a mistake in describing the previous image.

The image you sent me is a grid of white squares on a black background. This is a common geometric pattern that can be found in many contexts.

More specifically, this grid pattern is known as a Hermann grid illusion. The illusion is that the intersections of the grid appear darker than the squares themselves. This effect is caused by lateral inhibition in the human visual system.



Prompt Złudzenie Sandera



Wyświetl wersje odpowiedzi ▼ 🔊



The image you sent me is a black and white illustration of a face. It is made up of several geometric shapes, including triangles and circles.

Here's a more detailed description of the image:

- The face is looking straight ahead.
- The upper half of the face is comprised of four large triangles. Two triangles form the forehead and two triangles form the cheekbones.
- Two small circles create the eyes.
- A long, curved line creates the nose.
- The mouth is a thin line with a slight curve downwards at the corners, suggesting a neutral expression.
- There are no ears or hair depicted in this image.

This image is similar to the Kanizsa Triangle illusion, which uses our brain's ability to fill in shapes to create a complete image. In this case, the triangles and circles create the illusion of a face even though the complete shapes are not all there.