

„Studia Sandomierskie” 32 (2025), s. 283–290

ISSN 0208-7626

Ks. Łukasz Sadlocha

 <https://orcid.org/0000-0001-8093-3289>

lusad.lusad@gmail.com

Akademia Nauk Stosowanych im. Józefa Gołuchowskiego

 <https://ror.org/03dfz9238>

**Jeffrey Kane, *The Emergence of Mind. Where Technology Ends and We Begin*,
Palgrave Macmillan, London 2024, ss. 177
[recenzja]**

 <https://doi.org/10.15633/sts.3215>

Jeffrey Kane, *The Emergence of Mind. Where Technology Ends and We Begin*, Palgrave Macmillan, London 2024, ss. 177 [recenzja]

Abstrakt

Niniejszy tekst stanowi recenzję książki *The Emergence of Mind: Where Technology Ends and We Begin*, której autorem jest filozof Jeffrey Kane. Krytykuje on redukcjonistyczne podejście do AI, wątpiąc tym samym w możliwość zbudowania maszyny zdolnej do zajęcia miejsca człowieka.

Słowa kluczowe: sztuczna inteligencja, umysł, emergencja, antyredukcjonizm

Jeffrey Kane, *The Emergence of Mind. Where Technology Ends and We Begin*, Palgrave Macmillan, London 2024, 177 pages [review]

Abstract

This article is a review of the book *The Emergence of Mind: Where Technology Ends and We Begin*, written by the philosopher Jeffrey Kane. Kane criticizes the reductionist approach to AI, thus doubting the possibility of building a machine capable of replacing humans.

Keywords: artificial intelligence, mind, emergence, anti-reductionism

Ludzie współcześni doświadczają coraz większej integracji sztucznej inteligencji z codziennością. Nie brak głosów, że trudno już nawet mówić o inteligencji sztucznej, skoro podobno przeszła ona pomyślnie test Turinga¹. Niemniej wciąż dostrzega się granice istniejące między inteligencją sztuczną i ludzką. Do badaczy tych zagadnień przyłączył się także profesor filozofii edukacji Jeffrey Kane z Long Island University, zajmujący się doświadczalnymi podstawami ludzkiego myślenia. Swoje uwagi przedstawił w książce *The Emergence of Mind: Where Technology Ends and We Begin*, wydanej przez wydawnictwo Palgrave Macmillan w 2024 r.

Rozdział pierwszy (*Writing the Human Narrative*) pełni funkcję wstępu. Autor zaznacza, że celem książki jest zakwestionowanie obliczeniowego modelu myślenia leżącego u podstaw generatywnej AI (*artificial intelligence*) uważanej za ucieleśnienie ludzkiego umysłu i zaoferowanie alternatywnego podejścia, według którego to życie i świadomość są kluczowe dla tworzenia ludzkiej myśli. Idee, które przybierają formę w umyśle człowieka, są kształtowane przez doświadczenie, mają również swoje znaczenie i cel. Niestety wartości te są pomijane przez modele obliczeniowe. Należy zatem skonfrontować ów redukcjonizm obliczeniowych teorii umysłu opartych na statystycznej analizie języka z generatywnymi podstawami ludzkiego myślenia opartymi na dynamicznych zasadach pojawiających się w systemach złożonych.

Obliczeniowy model myślenia może doprowadzić do utracenia pewnych aspektów kondycji ludzkiej. Pożądane zaś jest wyjście poza redukcjonizm obliczeniowych teorii umysłu, aby zmierzyć się z generatywnymi podstawami ludzkiego myślenia. Alternatywna teoria umysłu jest relacyjna i opiera się na dynamicznych zasadach pojawiających się w systemach złożonych, a nie na statystycznych interakcjach.

W rozdziale drugim (*Where We Begin*) autor ukazuje podstawowy problem pojawiający się w przypadku generatywnej AI i podstawowych elementów ludzkiego doświadczenia. Jako przykład podaje on znaczenie słowa „niebieski”. AI nie ma doświadczenia tego koloru. Posługuje się tym wyrazem w danym znaczeniu tylko dlatego, że to człowiek nadał mu takie, a nie inne znaczenie w danym kontekście. Kane nawiązał przy tym do analogii badań nad wszechświatem do badań nad AI, zaproponowanej przez Yejin Choi. W fizycznym wszechświecie obserwuje się i bada ok. 5 proc. całej materii i energii, wiedząc zarazem, że na siłę grawitacji i prędkość ekspansji wpływa również pozostałe 95 proc. materii i energii, o której nic nie wiadomo. Podobnie jest w przypadku badań nad AI: naukowcy zajmują się jej niewielkim fragmentem, wiedząc, że istnieje jeszcze ogrom ciemnej materii inteligencji i języka.

Rozdział trzeci (*The Science of Mindlessness*) Kane rozpoczyna od ukazania związku między dwoma wymiarami ludzkiego istnienia, których doświadcza każdy człowiek: świadomości i cielesności. Obecna technologia umożliwia odtworzenie

¹ Od wielu lat wciąż trwa dyskusja na ten temat. Ostatnio odnotowany głos: C.R. Jones, B.K. Bergen, *Large Language Models Pass the Turing Test*, <https://arxiv.org/pdf/2503.23674> [dostęp: 31 III 2025 r.].

związku tych dwóch sfer, który mógł być zerwany np. w wyniku wypadku drogowego. Jest ona w stanie zrobić nawet więcej, gdyż może rozszerzyć przetwarzanie poznawcze (chipy Neuralink). Tym samym wydaje się, że ludzkie myślenie sprowadza się do przetwarzania informacji zgodnie z prawami fizyki i ukazuje umysł jako funkcję materii. Ale wtedy nie byłoby różnicy między AI a myśleniem człowieka. Myślenie nie byłoby cechą tylko ludzką. Modele algorytmiczne replikowałyby wzorce ludzkiego umysłu. Pozostaje tylko identyfikacja wzorców interakcji między obserwowalnymi obiektami, co przypomina dualizm kartezjuszowski.

Na myślenie człowieka – jak pisze autor – składają się jednak nie tylko informacje, ale i bezpośrednie doświadczenie odbierane przez neurony aferentne tworzące markery somatyczne, które następnie integrują się z innymi obszarami mózgu i ośrodkowym układem nerwowym, aby generować percepcje, uczucia i emocje. Osądy zatem powstają z procesów racjonalnych, jak też z bezpośredniego, subiektywnego doświadczenia wewnętrznego.

W rozdziale czwartym (*Where Computation Ends*) Kane wychodzi od koncepcji Kartezjusza. Medytacje francuskiego filozofa nad prawdą przebiegały w sferze koncepcyjnej, w oderwaniu od subiektywnego doświadczenia. Doprowadziło to do rozróżnienia dwóch światów, z których każdy wymagał własnej formy wiedzy. W świecie obiektywnym to rozum ustanawiał wszelki porządek, natomiast w świecie fizycznym – subiektywne wrażenia i interesy. Charles Darwin połączył te dwa światy, umieszczając ludzki umysł na poziomie biologicznym. Dla niego liczyły się już tylko relacje między środowiskiem a zachowaniami organizmu, dla których można było znaleźć wzorce matematyczne. Tym samym wyeliminowano subiektywność. Kolejną paradygmatyczną zmianę w badaniu ludzkiego umysłu zaproponował Noam Chomsky. Według niego umysł ludzki nie ogranicza się do środowiska fizycznego i obejmuje nieskończoną liczbę czynników psychicznych leżących poza środowiskiem, danym miejscem i czasem. Uwzględnienie takiej złożoności umożliwił dopiero rozwój technologii, zwłaszcza heurystyczne programy komputerowe i badania nad zrozumieniem ludzkich procesów heurystycznych i ich symulowanie (AI). Kolejnym badaczem był George Miller, według którego kluczem do tych dociekań powinny być struktury i wzorce zachowań językowych, czyli natura organizacji informacji, a nie jej reprezentacyjna (semantyczna) część ani relacja ze środowiskiem. Dzięki niemu dokonana się tzw. rewolucja poznawcza, a podstawowym kryterium dobrego modelu teoretycznego stała się obliczalność.

Metafora obliczeniowa umysłu wykluczyła wszystkie nieobliczeniowe stany wewnętrzne o znaczeniu naukowym w kontekście ludzkiego myślenia. Nastąpiło wywrócenie istniejącego porządku. Produkt, którym jest maszyna obliczeniowa, stał się prototypem umysłu ludzkiego. António Damásio, prowadząc badania nad funkcjonowaniem mózgu, odkrył, że subiektywnego doświadczenia i świadomości ludzkiej nie można zrozumieć obliczeniowo, nie są to też zjawiska nadprzyrodzone (jak dusza). Ponadto w codziennym użyciu informacja jest nośnikiem pewnej treści, natomiast w znaczeniu technicznym jest czysto funkcjonalna. Hubert Dreyfus odrzucił

rozumienie techniczne na rzecz codziennego w kontekście ludzkiego poznania. Dziś rozróżnienie to praktycznie zanikło w powszechnym użyciu.

Rozdział piąty (*Relational Science*) dotyczy ograniczeń, jakie niesie ze sobą funkcjonalne ujęcie inteligencji. Sztuczna inteligencja postępuje według określonego algorytmu, którego celem jest wygenerowanie wyników. Może rozwijać reguły przetwarzania (uczyć się, np. poprzez mechanizm sprzężenia zwrotnego), wydaje się nawet, że samodzielnie je wybiera. Mogą one również wykraczać poza możliwości projektantów, jednakże nie mogą generować nowych reguł przetwarzania poza parametrami wbudowanymi w ich zaprogramowaną architekturę. Nie są one w stanie porzucić swojego programu, choć człowiek może nie korzystać ze swojej intencji. To człowiek nadaje cel sztucznej inteligencji.

Ponadto trudno jest rozpatrywać mózg i jego działanie w oderwaniu od organizmu żywego, w którym wyewoluował i na którego potrzeby reaguje oraz odpowiada. Istnieje on w żywym podmiocie mającym zdolność subiektywnego doświadczania. To, że związek funkcji mózgowych z subiektywnym doświadczeniem nie jest kwestionowany, nie oznacza, że każde subiektywne doświadczenie da się sprowadzić do procesów neuronalnych. Modele obliczeniowe mózgu nie wystarczają do wyjaśnienia ludzkiej świadomości. Stwierdza to m.in. David Chalmers. Według niego obliczeniowy model umysłu zapewnia ogólne ramy dla wyjaśnienia procesów poznawczych. System można uznać za poznawczy, gdy znajdują się w nim informacje o algorytmicznym przetwarzaniu danych. Jednakże posiadanie takiego systemu nie oznacza jeszcze świadomego poznania związanego z subiektywnym doświadczeniem. Subiektywne doświadczenie może bowiem istnieć niezależnie od procesów obliczeniowych, i odwrotnie.

W rozdziale szóstym (*Life and Mind*) Kane rozprawia się z redukcjonizmem, który to ukazuje umysł jako coś nieobecnego we wszechświecie bądź w najlepszym razie nieistotnie sąsiadującego z nim. Wskazuje na relacyjność i na niepewność (prawdopodobieństwo) należące do istoty fizycznego wszechświata. Obiekty znajdujące się we wszechświecie składają się ze złożonych systemów i wykazują właściwości zależne od dynamicznej relacji pól materii i energii, w których istnieją. Z czasem pojawiają się nowe złożone systemy z własnymi systemami organizacji, w tym ludzie. Autor powołuje się m.in. na Philipa W. Andersona, według którego hipoteza redukcjonistyczna nie pociąga za sobą hipotezy konstruktywistycznej. Redukowanie wszystkiego do prostych elementów i praw nie oznacza możliwości rekonstrukcji wszechświata na ich podstawie. Dynamicznego zachowania dużych i złożonych agregatów cząstek nie da się zrozumieć za pomocą prostej ekstrapolacji właściwości kilku cząstek.

Kane opiera się również na teorii systemów (struktur) dyssypatywnych zainicjowaną przez belgijskiego fizyka Ilyę Prigogine'a. Otóż niektóre systemy fizyczne mają tendencję do naruszania zasady entropii. Organizują się, pochłaniając energię z otaczającego je środowiska. Wówczas jednak zwiększają własny porządek, a nie entropię. Tym samym stają się podstawą do pojawienia się nowych zasad organizacji

i nowych systemów dyssypatywnych. W przypadku znaczącego zakłócenia zasad organizacji taki system się rozpada.

Natura zatem jest zakorzeniona w niestabilności dynamicznej, a kolejne poziomy podlegają emergencji. Nowe poziomy są zależne od wszystkich zasad rządzących niższymi warstwami, a zarazem przejawiają własne emergentne zasady organizacyjne umożliwiające funkcjonowanie obiektów na ich poziomie. Stąd też brak jest przejść między prawami mechaniki kwantowej a prawami mechaniki klasycznej, między prawami biologii a prawami chemii, między prawami biologii a doświadczaniem wrażeń, a także między algorytmicznym przetwarzaniem informacji a świadomością. Redukcjonizm traktuje te braki jako luki dające się uzupełnić, co Kane uważa za złudzenie.

Rozdział siódmy (*Thinking as a Creative Act*) dotyczy życia biologicznego z perspektywy teorii systemów dyssypatywnych. Organizmy żywe, aby móc się utrzymać, potrzebują stałego dopływu energii z otaczającego ich środowiska. Stąd podstawową ich cechą jest inicjowanie działania. Oddziałują ze środowiskiem tak, by jak najmniejszym wysiłkiem wykorzystać jak największą ilość energii. Reagują na działania zewnętrzne, kierując się własnymi kryteriami. Wobec tego są selektywne w swoim zachowaniu i minimalizują błędy w swoich interakcjach. W konsekwencji takiego działania wyłoniło się również społeczeństwo ludzkie z własną organizacją, a w jego ramach powstają kolejne systemy z własnymi zasadami.

Również mózg powstał na drodze emergencji. Pozwala on bowiem wykonywać człowiekowi wiele funkcji niezakodowanych przez instynkt. Jego wyjątkową zdolnością jest tworzenie nowych i stopniowo coraz bardziej złożonych wzorców komunikacji między przedstawicielami własnego gatunku i szerszym środowiskiem. Z czasem język symboliczny umożliwił przekazywanie nie tylko obiektywnej wiedzy, ale i subiektywnych wrażeń – zarówno tego, co się wydarzyło, jak i tego, co się nie wydarzyło. Umożliwił on również rozwój bardziej złożonych i spójnych systemów organizacji społecznej. Ta nowa adaptacja do środowiska nie polega już na kolejnych pokoleniach powstających na drodze ewolucji genetycznej. Zmiany społeczne i językowe dokonują się w ciągu jednego życia. Z tej perspektywy sztuczna inteligencja nie jest powiązana z żadnym z ludzkich poziomów emergentnych.

Myślenie człowieka również powstało w wyniku emergencji ze wszystkich poziomów ludzkiego istnienia. Umysł organizuje sam siebie poprzez aktywność myślową, aby przekroczyć własne granice i znaleźć inny porządek poza nim. Pewne idee, będące elementami życia człowieka, nabierają coraz większej jedności i formy, przy czym stają się coraz bardziej podatne na konfrontację z innymi ideami. Systemy myśli wewnętrznie lub społecznie niespójne prawdopodobnie nie przetrwają. Tak jak ciało dąży do utrzymania swojej homeostatycznej integralności w kontekście czynników środowiskowych, tak też umysł dąży do utrzymania swojej systemowej integralności w kontekście wyzwań poznawczych. Człowiek żyje w świecie, świat zaś żyje w nim. Każdy człowiek poszukuje jedności, spójności samego siebie

i otaczającego go świata, a poszukiwane to jest źródłem inspiracji dla ludzkich myśli i idei, a zarazem głównym źródłem konfliktów i rozpaczy.

W rozdziale ósmym (*Emerging Mind*) autor podejmuje temat inteligencji. Fizyka nie może wyjaśnić aktów inteligencji, gdy inteligencję definiuje się jako zdolność wybierania spośród różnych opcji. Tak rozumiana inteligencja nie daje się również sprowadzić do programów obliczeniowych. Najważniejsze jest tu bowiem rozróżnienie postrzegającego świat i świata postrzeganego. Mechanizmy, za pomocą których dokonuje się wyborów, są drugorzędne. Fizyka nie wyjaśnia tego rozróżnienia, natomiast granicę tę tłumaczy biologia. Organizmy działają inteligentnie, utrzymując swoją homeostazę daleko od równowagi otaczającego je środowiska, działają celowo, lecz nie oznacza to automatycznie świadomości. Jako istoty świadome ludzie tworzą dodatkowo nowe wzorce oraz idee, dążą też do spójności w swoich systemach myślenia. Umysł wykracza poza własne doświadczenia i zainteresowania człowieka, język zaś umożliwia dzielenie się nimi z innymi. Tezy te Kane popiera konkretnymi badaniami i przykładami. Zaznacza zarazem, że obliczeniowe podejście do umysłu zuboży własne myślenie do samego przetwarzania informacji. Istota ludzka straci zdolność do rozumienia siebie jako czegoś więcej niż procesor informacji.

Autor recenzowanej książki w sposób popularnonaukowy ukazuje ograniczenia współczesnego modelu obliczeniowego jako modelu ludzkiego myślenia. Zamiast podejścia redukcijnego proponuje podejście emergencyjne. Książka posiada nieliczne przypisy, brakuje w niej również bibliografii. Przywoływani naukowcy są ogólnie znani. Wydaje się, że omawiana pozycja jest skierowana do osób dopiero rozpoczynających przygodę z filozoficznymi rozważaniami dotyczącymi ludzkiej i sztucznej inteligencji, ukazując im jedną z dróg.

Drogę tę można zauważyć, korzystając m.in. z klasyfikacji zaproponowanej przez Jarka Gryza, filozofa i informatyka zajmującego się kognitywistyką. Prześledził on dyskusję metodologiczną towarzyszącą sztucznej inteligencji od jej początku, wyróżnił trzy idee przyświecające wszystkim, którzy usiłują zbudować systemy percepcji. Jedni uważają, że celem jest zbudowanie maszyny imitującej zachowanie (a więc i myślenie) człowieka (ang. *human-level AI*). Jednakże chociaż do tego nurtu należą nestorzy AI, to obecni jego reprezentanci stanowią mniejszość. Drudzy apelują o odrzucenie ludzkiego zachowania i myślenia jako wzorca dla AI. Uważają, że warto zrozumieć ludzkie myślenie, jednak celem jest zbudowanie AI przekraczającej ludzkie zdolności umysłowe w jakiś użyteczny sposób. Ostatnia, najlicniejsza grupa, postuluje całkowite porzucenie idei maszyny myślącej, proponując skupienie się na „podmiotach racjonalnych” (ang. *rational agents*), które maksymalizują zadaną miarę wydajności. Sugerują nawet zmianę nazwy dziedziny ze „sztucznej inteligencji” na „racjonalność obliczeniową”².

² J. Gryz, *Gdzie jesteś, HAL?*, „Przegląd Filozoficzny – Nowa Seria” 22 (2013), nr 2 (86), s. 181.

Zgodnie z podejściem emergentystycznym Kane nie neguje prac nad „podmiotami racjonalnymi” ani pozaludzkimi AI. Wątpi natomiast w możliwość zbudowania maszyny imitującej człowieka (w tym AI), a tym samym zajmującej niejako jego miejsce.

Bibliografia

Opracowania

Jones C.R., Bergen B.K., *Large Language Models Pass the Turing Test*, <https://arxiv.org/pdf/2503.23674> [dostęp: 31 III 2025 r.].

Gryz J., *Gdzie jesteś, HAL?*, „Przegląd Filozoficzny – Nowa Seria” 22 (2013), nr 2 (86), s. 167–184.

Kane J., *The Emergence of Mind. Where Technology Ends and We Begin*, Palgrave Macmillan, London 2024.