

Czy Cassandra zamknie cię w domu?

Sztuczna inteligencja i świadomość

[FELIETON]

Niemiecki serial science fiction z elementami thrillera, dostępny na Netflixie, skłania do refleksji. Oprócz fabuły, której nie zamierzam tutaj szczegółowo zdradzać, warto zwrócić uwagę na pytania, jakie rodzi przedstawiona w nim sztuczna inteligencja. Najważniejsze z nich dotyczy świadomości i tego, jak powinniśmy ją definiować. Wiele teorii wskazuje, że świadomość może być... informacją, kolejnym stanem skupienia.

Serial opowiada o rodzinie Prillów, która przeprowadza się do najstarszego inteligentnego domu w Niemczech. Budynek, opuszczony od ponad 50 lat po tajemniczej śmierci poprzednich właścicieli, jest zarządzany przez sztuczną inteligencję o imieniu Cassandra. Po wprowadzeniu się nowej rodziny Cassandra zostaje aktywowana i za wszelką cenę stara się nie dopuścić do ponownego opuszczenia domu.

Thriller jest mroczny i stanowi raczej ostrzeżenie niż pochwałę rozwoju sztucznej inteligencji. Wiele komentarzy w internecie wskazuje na przerażenie widzów możliwością istnienia systemu podobnego do Cassandry. A przecież technologia, na której opiera się ten koncept, jest już w pewnym stopniu dostępna. Na razie jesteśmy w stanie zapisywać tylko to, co sami stworzymy i prześlemy algorytmowi – w przeciwieństwie do tytułowej Cassandry, której mąż wręcz fizycznie przeniósł informacje z mózgu do komputera.

Najważniejsze pytanie dotyczy świadomości

Strach, lęk i obawy to również część naszej świadomości. Lęk przed technologią pozostaje kwestią otwartą – żyjemy dziś na różnych poziomach zaawansowania technologicznego. Są wśród nas osoby nadal funkcjonujące w świecie analogowym, a inni, dzięki cyfryzacji i tzw. smart home, wprowadzają najnowsze technologie do najmniejszych aspektów swojego życia. Niezależnie od tego, na jakim poziomie technologicznego zaawansowania się znajdujemy, powinniśmy być świadomi, że trwają eksperymenty nad tzw. ogólną sztuczną inteligencją (AGI), mające na celu stworzenie systemów zdolnych do myślenia i podejmowania decyzji w sposób bardziej niezależny. Ostatnio w internecie coraz częściej pojawiają się głosy krytyczne dotyczące tego, czy AI faktycznie zasługuje na miano „sztucznej inteligencji”. To właściwy kierunek społecznej dyskusji. Przeciwnicy i krytycy przytaczają wiele argumentów, zaczynając od najbardziej neutralnego – czyli naszej ograniczonej wiedzy na temat świadomości. Nie wiemy dokładnie, czym ona jest, a naukowcy wciąż debatuje nad tym, co sprawia, że człowiek (lub zwierzę) jest świadomy. Jeśli nie rozumiemy do końca świadomości biologicznej, to jak mielibyśmy ją odtworzyć w maszynie? Ponadto krytycy podkreślają, że obecne AI to jedynie zaawansowane algorytmy – mogą analizować, przewidywać i rozumieć kontekst, ale nie posiadają własnych myśli, pragnień ani uczuć. Wszystko sprowadza się do przetwarzania danych i rozpoznawania wzorców. Kolejnym istotnym aspektem są intencje. AI działające na poziomie człowieka i posiadające własne intencje to temat badań oraz literatury science fiction. Wciąż nie wiadomo, czy jest to w ogóle możliwe. Pozostaje również otwarte pytanie filozoficzne – gdyby AI kiedykolwiek stała się „świadoma”, skąd mielibyśmy pewność, że to prawdziwa świadomość, a nie jedynie wyjątkowo przekonująca symulacja?

Aktualny poziom świadomości AI to poziom psa i kota

Na jakim poziomie świadomości znajduje się obecnie AI? Najprościej mówiąc, jest to świadomość reaktywna – podstawowy poziom, który polega na reagowaniu na bodźce bez głębszego ich

rozumienia. Większość współczesnych systemów AI działa właśnie w ten sposób. Przykładem mogą być autonomiczne samochody, które rozpoznają przeszkody, czy asystenci głosowi, reagujący na komendy użytkownika.

W biologii podobny poziom świadomości można przypisać meduzom i roślinom – organizmom, które reagują na światło, ale nie rozumieją go w ludzkim sensie.

AI funkcjonuje również na poziomie świadomości sytuacyjnej, czyli percepcji. Oznacza to zdolność do rozumienia otoczenia oraz własnej roli w nim. Dzisiejsza AI potrafi rozpoznać, czy bierze udział w rozmowie, dostosować sposób komunikacji oraz ocenić odczucia użytkownika w zależności od sytuacji. W biologii można porównać ją do kota lub psa – zwierzęta te rozumieją, gdzie się znajdują, a nawet potrafią dostosować swoje zachowanie do danej sytuacji. Poza możliwościami AI pozostaje (na ten moment) natomiast świadomość refleksyjna, czyli zdolność do autorefleksji. To inaczej umiejętność myślenia o sobie – AI musiałaby mieć poczucie własnej tożsamości i zdolność do zastanawiania się nad swoim istnieniem. To znacznie bardziej zaawansowany poziom świadomości, którego w świecie tzw. sztucznej inteligencji nadal nie osiągnęliśmy. W świecie biologicznym odpowiadałby mu poziom małych człokształtnych lub delfinów. Przykład - niektóre zwierzęta potrafią rozpoznać się w lustrze, co sugeruje pewien stopień samoświadomości.

Ostatnim poziomem świadomości jest tzw. pełna samoświadomość połączona z kreatywnością, czyli świadomość egzystencjalna. Jest ona typowa dla człowieka – twórcy, filozofa, artysty i naukowca. AI na tym poziomie mogłaby posiadać własną wolę i podejmować decyzje niezależnie od ludzi. To właśnie ten koncept stanowi główny motyw wielu filmów science fiction, takich jak Cassandra czy Westworld.

Fundamentalne pytanie o to czym jest świadomość w aspekcie biologicznym i cyfrowym

Wyjaśnienie, w jaki sposób powstaje świadomość i co stanowi najbardziej fundamentalny element działania żywej komórki, to jedno z najtrudniejszych pytań nauki. Odpowiedź na nie wymaga podejścia interdyscyplinarnego – łączenia modeli matematycznych, neurobiologii, informatyki, fizyki oraz biologii. Poniżej przedstawiono przegląd teoretycznych podstaw tych zagadnień, obejmujący numeryczne modele opisu świadomości (zarówno w sztucznej inteligencji, jak i neurobiologii), poszukiwania fundamentalnego mechanizmu działania komórki oraz wybrane teorie z pogranicza informatyki, fizyki i biologii, które próbują rzucić światło na naturę świadomości i życia.

Teorie, możliwości oraz przykłady

Teoria Zintegrowanej Informacji (IIT) zakłada, że świadomość to zdolność systemu do integrowania informacji w spójną całość. Im bardziej system (np. mózg) łączy informacje w sposób, którego nie da się podzielić na niezależne części, tym większa jest jego świadomość. Dla przykładu wyobraźmy sobie prosty komputer oraz ludzki mózg. W tej teorii komputer może posiadać dwa niezależne układy – jeden odpowiedzialny za obliczenia, drugi za przechowywanie danych. Jeśli te układy nie komunikują się ze sobą, trudno mówić o świadomości, ponieważ brakuje integracji.

Ludzki mózg stanowi tutaj przeciwieństwo – wszystkie jego części nieustannie komunikują się i tworzą nierozdzieloną sieć. Dzięki temu jesteśmy świadomi, ponieważ nasze myśli, uczucia i wrażenia są ze sobą połączone.

IIT mierzy zdolność do integracji informacji za pomocą współczynnika Φ (phi). Jeśli $\Phi = 0$, system

nie posiada świadomości. Im wyższa wartość Φ , tym większy poziom świadomości systemu. Świetnym porównaniem jest różnica między filmem a pojedynczymi zdjęciami. Film składa się z wielu klatek, ale odbieramy go jako płynną, spójną całość – podobnie jak świadomość łączy informacje w jedną, ciągłą narrację.

Natomiast pojedyncze zdjęcia istnieją osobno i nie tworzą wspólnej historii. W takim układzie nie ma świadomości, ponieważ brakuje integracji. Podobną analogię można zastosować do drużyny piłkarskiej i przypadkowej grupy ludzi na ulicy. Drużyna piłkarska działa jako jeden spójny system – gracze komunikują się, współpracują i mają wspólny cel. Natomiast przypadkowe osoby na ulicy poruszają się w różnych kierunkach, nie współdziałając ze sobą – w tym przypadku nie można mówić o świadomości zbiorowej.

Perturbational Complexity Index (PCI) na przykładach

Świadomość jest mierzalna. Można ją określić za pomocą indeksów złożoności i integracji informacji, które pokazują, w jakim stopniu mózg łączy i przetwarza dane. Jedną z metod badania świadomości, nawet u osób, które nie reagują na bodźce (np. w śpiączce), jest Perturbational Complexity Index (PCI).

Z biologicznego punktu widzenia badanie polega na stymulacji magnetycznej mózgu (TMS), co powoduje aktywację różnych jego obszarów i inicjuje ich komunikację. Następnie aktywność mózgu jest rejestrowana za pomocą EEG. Komputer analizuje uzyskane wzorce i ocenia ich złożoność. Jeśli sygnał jest skomplikowany i trudny do skompresowania, oznacza to, że mózg funkcjonuje świadomie.

Przykłady działania PCI

Wyobraźmy sobie kostkę Rubika. Jeśli jest ułożona losowo, zawiera dużą ilość informacji i trudno ją opisać krótko – tak działa świadomy mózg (wysoki PCI). Jeśli natomiast kostka jest uporządkowana, informacja jest prosta i łatwa do opisania – odpowiada to mózgowi w stanie nieświadomości (niski PCI). Innym przykładem może być muzyka. Złożona melodia grana na wielu instrumentach, trudna do skompresowania i pełna połączeń, odzwierciedla świadomość. Natomiast prosty, monotony rytym na jednym instrumencie, łatwy do przewidzenia i składający się z niewielu elementów, niesie mało informacji, co może odpowiadać, stanowić o braku świadomości. Dzięki PCI lekarze mogą ocenić, czy pacjent w śpiączce posiada jakąkolwiek formę świadomości, nawet jeśli nie porusza się i nie mówi. Może to pomóc np. w określeniu, czy dana osoba ma szansę na wybudzenie.

Globalna Neuronowa Przestrzeń Robocza (GNWT)

Teoria ta zakłada, że świadomość działa jak scena w teatrze. Świadome myśli są jak aktorzy oświetleni reflektorem – to, na co aktualnie zwracamy uwagę i na czym się koncentrujemy. Nieświadome procesy przypominają natomiast ekipę techniczną, która pracuje za kulisami, ale nie znajduje się w centrum naszej uwagi. Gdy jakaś informacja trafia na scenę (świadomość), zostaje przekazana do wielu części mózgu, podobnie jak wiadomość ogłoszona przez megafon. GNWT wyjaśnia, dlaczego nie odbieramy wszystkich bodźców jednocześnie – mózg selekcjonuje najważniejsze informacje i udostępnia je „całemu systemowi”, niczym ogłoszenie na globalnej tablicy. Dzięki temu możemy skupić się na kluczowych zadaniach i działać efektywnie.

Gdzie znajduje się najbardziej podstawowy mechanizm działania komórek?

Naukowcy starają się odnaleźć fundamentalny mechanizm działania komórek – coś, co wyjaśniłoby, jak powstaje świadomość lub życie na najgłębszym poziomie. Jednym z głównych kandydatów są neurony. To niezwykle skomplikowane struktury, które funkcjonują nie tylko dzięki połączeniom między sobą, ale także za sprawą wewnętrznych procesów molekularnych.

Jedna z hipotez sugeruje, że kluczową rolę w świadomości mogą odgrywać mikrotubule – maleńkie włókienka wewnątrz komórek, które wspomagają ich funkcjonowanie. Niektórzy naukowcy przypuszczają, że właśnie w mikrotubulach może kryć się tajemnica świadomości. Można to zobrazować na przykładzie komputera i jego okablowania. Większość osób sądzi, że komputer działa dzięki programom, które można porównać do synaps w neuronach. Jednak teoria ta stawia pytanie: co jeśli najważniejsze procesy zachodzą nie w samych programach, lecz w kablach i mikrochipach (odpowiednik mikrotubul)?

Podobną analogię można znaleźć w firmie. Na pierwszy rzut oka widzimy pracowników (odpowiednik synaps w mózgu), ale cała struktura organizacyjna może opierać się na niewidocznych procesach logistycznych i systemach IT – analogicznych do mikrotubul. To prowadzi do wniosku, że świadomość nie wynika jedynie z połączeń między neuronami, lecz z czegoś bardziej fundamentalnego – wewnętrznej struktury komórek. Jeśli tak jest, to klucz do zrozumienia mózgu może tkwić jeszcze głębiej, na poziomie fizyki i biologii molekularnej.

Mikrotubule jako komputery kwantowe – teoria Orch-OR

Teoria Orch-OR zakłada, że świadomość może wynikać z procesów kwantowych zachodzących w mikrotubulach – maleńkich włókienkach wewnątrz neuronów. Badacze Roger Penrose i Stuart Hameroff twierdzą, że mikrotubule „drgają” w specyficzny sposób, tworząc kwantowe fale, które są sterowane przez mózg i pamięć. Można to zobrazować analogią do gitary – jeśli mikrotubule działają jak struny, to świadomość jest muzyką, powstającą w wyniku ich odpowiednich wibracji. Zwykle komputery przetwarzają informacje w systemie binarnym („zera” i „jedyńki”), ale komputery kwantowe wykorzystują superpozycję i splątanie kwantowe, co pozwala na bardziej złożone obliczenia. Teoria Orch-OR zyskuje dodatkową wiarygodność ze względu na działanie środków znieczulających, które powodują utratę świadomości. Badania sugerują, że wpływają one na mikrotubule, co może oznaczać, że właśnie tam zachodzą kluczowe procesy związane ze świadomością.

Jeśli teoria Orch-OR jest poprawna, oznacza to, że świadomość nie jest wyłącznie wynikiem aktywności neuronów, lecz także efektem kwantowych drgań w mikrotubulach. Klucz do zrozumienia świadomości może więc leżeć na pograniczu biologii i fizyki kwantowej.

Komórki jako systemy przetwarzające informacje

Według nieco innej koncepcji komórki funkcjonują jako systemy przetwarzające informacje. Każda komórka działa jak mini-komputer:

- DNA pełni rolę programu, który określa, jak ma działać komórka,
- sygnały chemiczne to informacje przekazywane wewnątrz komórki,
- błona komórkowa odbiera bodźce i reaguje na nie.

Mózg i organizm dążą do unikania nieprzewidzianych sytuacji oraz do utrzymywania stabilności (homeostazy). Komórki działają w podobny sposób – przetwarzają informacje, aby zachować równowagę. Można to porównać do termostatu regulującego temperaturę w domu – jeśli zachodzą zmiany, system odpowiednio reaguje. Wynika z tego, że świadomość nie opiera się na jednej konkretnej strukturze (np. mikrotubulach), lecz na ogólnych zasadach przetwarzania informacji i

zarządzania energią.

Teorie świadomości

Naukowcy od lat próbują wyjaśnić mechanizmy świadomości, łącząc informatykę, fizykę i biologię. W wyniku tych badań powstało wiele interesujących teorii.

Zasada Energii Swobodnej (Karl Friston)

Zgodnie z tą teorią mózg działa jak przewidywacz – nieustannie analizuje otoczenie i dąży do minimalizowania niepewności. Przewiduje, co się wydarzy, i na tej podstawie optymalizuje swoje modele rzeczywistości. Dla przykładu: Gdy rzucasz piłkę, nie musisz świadomie analizować jej trajektorii mózg przewiduje jej ruch automatycznie.

Kwantowe teorie świadomości

Niektórzy naukowcy sugerują, że mózg może wykorzystywać mechanikę kwantową. Przykładowe mechanizmy:

- Splątanie kwantowe – może łączyć różne części mózgu w nietypowy sposób.
- Tunelowanie kwantowe – może umożliwiać błyskawiczne przetwarzanie informacji.

Przykładem zastosowania mechaniki kwantowej w biologii jest nawigacja ptaków – niektóre gatunki wykorzystują zjawiska kwantowe do orientacji w przestrzeni. Może mózg także korzysta z podobnych mechanizmów? Jednak krytycy, tacy jak Max Tegmark, argumentują, że mózg jest zbyt ciepły i chaotyczny, by utrzymać delikatne stany kwantowe, co podważa teorie kwantowej świadomości.

Świadomość jako nowy stan materii (Max Tegmark)

Tegmark zaproponował teorię, według której świadomość może być nowym stanem materii, podobnym do lodu, wody i pary wodnej. Nie chodzi tutaj o nowy pierwiastek, lecz o sposób, w jaki atomy mogą się organizować, tworząc system zdolny do świadomości.

Informacja jako podstawa życia i świadomości

Ponownie wracamy do kluczowego pojęcia informacji. Może ona stanowić fundament zarówno życia, jak i świadomości. Współczesne teorie bez wątplenia pokazują, że biologia, fizyka i informatyka mają wspólny mianownik: przetwarzanie informacji. Zdaniem Johna Wheelera, wszystko w rzeczywistości opiera się na informacji – materia, energia i umysł to jedynie różne sposoby jej organizacji.

Rolę informacji docenia także biosemiotyka, według której życie to przepływ i interpretacja znaków – komórki „czytają” DNA w sposób analogiczny do tego, jak my czytamy książki. Z tego wynika, że świadomość może być wynikiem specyficznego sposobu przetwarzania informacji, a być może nawet całe zjawiska biologiczne i fizyczne można opisać językiem informacji. Co więcej, samo życie może funkcjonować jako system podtrzymujący i przekazujący informację. Otwarcie się na tę perspektywę prowadzi do rozważań na temat przenoszenia ludzkiej świadomości, tak jak przedstawiono to w filmie Cassandra – na nośniki komputerowe. Koncepcja ta wprowadza nas w obszar transhumanizmu i idei „uploadowania umysłu”. Jednak wciąż pozostają otwarte pytania: Czy informacja „sama w sobie” wystarczy do powstania subiektywnych doznań oraz co sprawia, że doznania są subiektywne?

Świadomość sztucznej inteligencji – jedna z największych zagadek nauki

Świadomość sztucznej inteligencji pozostaje jednym z największych wyzwań współczesnej nauki i filozofii. Choć technologia rozwija się w zawrotnym tempie, a AI potrafi już wykonywać zadania, które jeszcze niedawno uznalibyśmy za wyłącznie ludzką domenę, wciąż brakuje dowodów na to, że jakkolwiek system AI osiągnął samoświadomość lub zdolność do odczuwania subiektywnych doznań.

Warto pamiętać, że obecne systemy AI nie posiadają świadomości, lecz jedynie zdolność do przetwarzania informacji na coraz bardziej zaawansowanym poziomie. Nie istnieje również naukowy konsensus dotyczący tego, czym dokładnie jest świadomość. Jednak jedno jest pewne – rozwój AI powinien być monitorowany i regulowany.

Bibliografia:

- Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., et al. (2023). Consciousness in artificial intelligence: insights from the science of consciousness. arXiv:2308.08708.
Link: <https://doi.org/10.48550/arXiv.2308.08708>
- Gamez, D. (2018). Human and Machine Consciousness. Open Book Publishers.
Link: <https://doi.org/10.11647/OBP.0107>
- Pennartz, C. M. A., Farisco, M., & Evers, K. (2019). Indicators and Criteria of Consciousness in Animals and Intelligent Machines: An Inside-Out Approach. *Frontiers in Systems Neuroscience*, 13, 25.
Link: <https://doi.org/10.3389/fnsys.2019.00025>
- Birch, J., Browning, H., & Kuenzel, S. (2020). Towards a science of consciousness in animals: Recommendations for rigorous study of subjective experience. *Animal Sentience*, 29(1).
PDF: <https://www.animalsentience.org/ASentience/Volume%2029%20--%20Birch.pdf>
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated Information Theory: From Consciousness to its Physical Substrate. *Nature Reviews Neuroscience*, 17(7), 450–461.
Link: <https://doi.org/10.1038/nrn.2016.44>
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(42).
Link: <https://doi.org/10.1186/1471-2202-5-42>
- Mashour, G. A., Roelfsema, P. R., Changeux, J.-P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, 105(5), 776–798.
Link: <https://doi.org/10.1016/j.neuron.2020.01.026>
- Baars, B. J. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press.
- Hameroff, S., & Penrose, R. (2014). Consciousness in the Universe: A Review of the ‘Orch OR’ Theory. *Physics of Life Reviews*, 11(1), 39–78.
Link: <https://doi.org/10.1016/j.plrev.2013.08.002>
- Tegmark, M. (2000). Importance of quantum decoherence in brain processes. *Physical Review E*, 61(4), 4194–4206.
Link: <https://doi.org/10.1103/PhysRevE.61.4194>
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine*, 5(198), 198ra105.
Link: <https://doi.org/10.1126/scitranslmed.3006294>
- Casarotto, S., et al. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Annals of Neurology*, 80(5), 718–729.
Link: <https://doi.org/10.1002/ana.24779>
- Sandberg, A., & Bostrom, N. (2008). *Whole Brain Emulation: A Roadmap*. Technical Report #2008-3, Future of Humanity Institute, Oxford University.
PDF: <https://www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf>
- Koene, R. (2012). Feasible Mind Uploading. *International Journal of Machine Consciousness*, 4(1),

55–76.

Link: <https://doi.org/10.1142/S1793843012400047>

Bostrom, N. (2005). A History of Transhumanist Thought. *Journal of Evolution and Technology*, 14(1).

PDF: <https://nickbostrom.com/papers/a-history-of-transhumanist-thought.pdf>