

## **Sztuczna inteligencja**

Autor tekstu: **Bernard Korzeniewski**

O mawiając kwestię natury oraz fizycznych (neurofizjologicznych) korelatów ludzkiej psychiki i (samo)świadomości, nie sposób nie wspomnieć chociażby pokrótce o problemie sztucznej inteligencji — AI (od angielskiego *Artificial Intelligence*). Problem ten dotyczy warunków, jakie musiałyby zostać spełnione, aby urządzenia skonstruowane przez człowieka, na przykład komputery, roboty czy sieci neuropodobne, posiadały inteligencję i psychikę. Określenie tych warunków rzuciłoby dużo światła na istotę obu interesujących nas fenomenów.

Oczywiście już samo stwierdzenie, że, przynajmniej w zasadzie [1], możliwe jest stworzenie sztucznych urządzeń obdarzonych subiektywną sferą odczuć psychicznych, posiada daleko idące konotacje światopoglądowe. Zakłada ono bowiem milcząco, że: 1. do wyłonienia się świadomości nie jest potrzebny żaden jakościowo odmienny byt (w rozumieniu filozoficznym), na przykład duch, oraz 2. fizycznym korelatem świadomości nie jest jakiś konkretny rodzaj tworzywa (np. neurony zbudowane z białek, lipidów itd.), lecz raczej pewien szczególny rodzaj funkcjonalnej organizacji sieci elementów zbudowanych z praktycznie dowolnej substancji (metal, krzem, białko), lecz pozostających ze sobą w odpowiednich związkach formalnych (przede wszystkim informacyjnych i cybernetycznych). Jak już się zapewne Czytelnik zdołał doskonale zorientować, należę do gorących zwolenników takiej właśnie opcji. Nie brak jednakże głosów, iż z zasady niemożliwe jest stworzenie *deus ex machina*, czyli sztucznej inteligencji. Zanim przedstawię jeden z argumentów na rzecz tej tezy, spróbuję najpierw zdefiniować, co w istocie rzeczy należy (moim zdaniem) rozumieć przez „sztuczną inteligencję”.

„Naturalna inteligencja”, czyli ludzka sfera mentalna, posiada wiele aspektów czy też komponent. Są to, między innymi, potocznie rozumiana inteligencja, psychika, (samo)świadomość, subiektywizm, myślenie, odczucia emocjonalne, osobowość. Powstaje od razu problem, czy wszystkie te aspekty są z konieczności wzajemnie sobą uwarunkowane (u człowieka występują one razem), czy też, przeciwnie, mogą być rozdzielone i istnieć niezależnie, a jeśli tak, to które z nich mogą warunkować (w największym stopniu) zaistnienie sztucznej inteligencji. Przecież wiele współczesnych komputerów i robotów (a także wiele zwierząt) zachowuje się, wedle potocznych kanonów, inteligentnie, może nawet „myśli”, chociaż raczej odmówilibyśmy im samoświadomości, osobowości i sfery subiektywnych odczuć emocjonalnych. Jednakże, moim zdaniem, „prawdziwa” inteligencja i myślenie polega przede wszystkim na stawianiu samemu sobie problemów i celów, a nie li tylko na rozwiązywaniu zadań narzuconych z zewnątrz, przez człowieka lub ewolucję, jak to ma miejsce, odpowiednio, w przypadku komputerów i zwierząt. Dlatego też sądzę, iż wszystkie wymienione powyżej aspekty subiektywnej sfery mentalnej wzajemnie się współwarunkują i nie mogą istnieć jedno bez drugich. Na przykład autonomiczne myślenie prowadzi do realizacji jakiegoś celu, co wiąże się z pozytywnymi odczuciami emocjonalnymi przeżywanymi przez obdarzony (samo)świadomością i posiadający osobowość podmiot. Idea wytworzenia „czystego rozumu”, pozbawionego motywacji (a więc emocji) i nieświadomego swego własnego istnienia stanowi dla mnie utopijną mrzonkę, niemożliwą do urzeczywistnienia z przyczyn zasadniczych i świadcząca o całkowitym niezrozumieniu istoty podłoża ludzkiej psychiki. Dlatego wyrażenie „sztuczna inteligencja”, pośrednio sugerujące możliwość istnienia takiego rozumu, uważam za szczególnie niefortunne (no cóż, przyjęło się, i dlatego zmuszony jestem je stosować). Wedle mojej opinii warunkiem niezbędnym dla zaistnienia prawdziwej (psychicznej) inteligencji jest pojawienie się (samo)świadomości. Diagnoza ta z pewnością nie stanowi zaskoczenia dla Czytelników, którzy zapoznali się z poprzednim rozdziałem. [2]

Ustaliliśmy już w przybliżeniu, co rozumiemy pod nazwą sztucznej inteligencji, musimy teraz określić kryteria pozwalające rozstrzygać o tym, czy dane urządzenie obdarzone jest psychiką, czy też nie. Powszechnie znany jest test Turinga, polegający na tym, iż wymieniamy (np. zapisane na kartkach papieru) pytania i odpowiedzi z „rozmówcą”, którego nie widzimy, bo znajduje się w innym pokoju. Nasza rola polega na stwierdzeniu, czy mamy do czynienia z człowiekiem, czy z maszyną. Jeśli nie potrafimy jednoznacznie stwierdzić, czy nasz „rozmówca” jest człowiekiem, czy komputerem, a okaże się, że był to komputer, będzie to oznaczało, iż przeszedł on test Turinga, a zatem obdarzony jest psychiką i świadomością. Na teście Turinga oparty został opracowany przez Searle’a eksperyment myślowy, mający jakoby zdyskredytować samą możliwość skonstruowania

sztucznej inteligencji, a nazwany „chińskim pokojem”. W pokoju tym zamiast komputera znajduje się grupa Chińczyków, którzy nie znają języka angielskiego, ale otrzymują karteczki z pytaniami zadawanymi w tym właśnie języku. Chińczycy dostarczają odpowiedzi na owe pytania (także na karteczkach), operując zestawem instrukcji w języku chińskim (który to zestaw odpowiada formalnie algorytmowi pewnego programu komputerowego), określających zespół czysto mechanicznych reguł dotyczących przekształcania sekwencji angielskich wyrazów w pytaniach w sekwencję angielskich wyrazów w odpowiedziach. Jeśli taki zespół reguł byłby dostatecznie bogaty, grupa Chińczyków przeszłaby test Turinga (zostałaby uznana za układ świadomy), choć w najmniejszym stopniu nie rozumiałaby ani zadawanych pytań, ani udzielanych odpowiedzi. A zatem komputer, którego zadanie także polega jedynie na mechanicznym przekształcaniu danych wedle z góry dostarczonych zasad, nie może posiadać świadomości.

Uważam, że pozorny „paradoks” chińskiego pokoju Searle’a jest mylący i w istocie swej niemądry. Po pierwsze, trzeba by najpierw wykazać, że test Turinga stanowi rzeczywiście wiarygodne kryterium posiadania przez badany system psychiki. [3] Po drugie, wcale nie jest (dla mnie przynajmniej) oczywiste, czy pokój pełen Chińczyków przeszedłby test Turinga. Po trzecie, zastąpienie mikroprocesorów w komputerze Chińczykami w pokoju jest chwytem tyleż bałamutnym, co heurystycznie chybnym. Fakt, że pojedynczy Chińczyk, będący przecież istotą świadomą, nie rozumie angielskiego, o niczym nie świadczy — przecież oczywiste jest, że pojedynczy mikroprocesor komputera (odpowiednik Chińczyka) także nie rozumie operacji wykonywanych w programie komputerowym. Tym, co może stanowić podłoże świadomości, jest dopiero wielce i w specjalny sposób złożony system mikroprocesorów. Analogia do ludzkiego mózgu nasuwa się tu w sposób oczywisty: dopiero cały system, a nie któryś z jego elementów (neuronów), może być świadomy. Searle wcale nie pokazał, że system: Chińczycy + dostarczony im zespół reguł nie jest w stanie *jako całość* (choć może się to wydać dziwne) być świadomy i rozumieć angielskiego. Psychika stanowi zjawisko emergentne i jako takie domaga się przyznania obiektywnej egzystencji fenomenowi złożoności. Przed tym jednakże wiele nauk przyrodniczych, które rozwinęły się na bazie zbyt prymitywnie rozumianego redukcjonizmu, nadal usilnie się wzbrania.

Innym błędem w rozumowaniu Searle’a jest przypisanie sztucznej inteligencji zasady działania współczesnego, digitalnego komputera, realizującego pewien ściśle zdeterminowany algorytm programu oraz operującego dokładnie określonym zestawem dyskretnych symboli i reguł rządzących ich potencjalnymi relacjami (czyli składnią). Analogią takiego właśnie komputera jest przecież pokój z Chińczykami. Dla tej też przyczyny (a nie z powodów implikowanych jakoby przez omawiany eksperyment myślowy) uważam, że zespół reguł translacyjnych obsługiwany przez Chińczyków, mikroprocesory czy odpowiednio wytresowane krewetki nie mogłyby stać się świadome. Skoro tak, to czego brakuje współczesnym komputerom, i co musiałyby osiąść przyszłe elektroniczne układy przetwarzające informację, aby wejść w posiadanie psychiki?

Pominę, oczywistą chyba, kwestię wielkości i stopnia złożoności całego układu. Przechodząc do następnego punktu, zacznę od stwierdzenia, że „komputer” taki powinien działać, w przeciwieństwie do większości jego współczesnych krewniaków, tak jak sieć neuronalna: w sposób analogowy, przybliżony i niealgorytmiczny. Nie powinien być oparty na dyskretniej logice zero-jedynkowej. Przetwarzanie informacji powinno się odbywać w sposób równoległy, a sam system przetwarzający powinien móc ulegać modyfikacji i ewolucji (uczeniu się), tak jak współczesne sieci neuropodobne. Poszczególne elementy systemu nie powinny działać w sposób całkowicie z góry określony, lecz uczyć się swoich funkcji poprzez interakcję z innymi elementami systemu, a także na bazie bodźców napływających do systemu z zewnątrz. O ile więc obecne komputery „myślą” w warstwie językowej (dotyczyło to także, przynajmniej wedle jego własnego mniemania, Ludwiga Wittgensteina), poprzez operację na określonych dyskretnych symbolach porządkowanych za pomocą odpowiednich reguł gramatycznych, to świadomy komputer działałby w oparciu o leżącą głębiej warstwę pojęciową (nieobecną w komputerach współczesnych), gdzie poszczególne elementy znaczyłyby przez konotację, tak jak neurony w sieci neuronalnej. W „sieci pojęciowej” takiego komputera język byłby tworem wtórnym, wysokiego rzędu, nieautonomicznym, bo semantycznie zależnym od rzeczowej sieci.

Obdarzony psychiką komputer musiałby mieć jakiś odpowiednik centrum motywacyjnego, „siedliska popędów”, a także połączonego z nim ośrodka nagrody/kary w ludzkim mózgu. Ośrodek ten utrzymywałby te połączenia w sieci elementów neuropodobnych, które sprzyjałyby realizacji jakichś autonomicznych celów odpowiadających popędom (u organizmów żywych są one narzucone przez logikę ewolucji biologicznej), np. popędowi poznawczemu, samozachowawczemu itp. Całość stanowiłaby podstawę subiektywnej sfery odczuć emocjonalnych. Konieczna wydaje się także spontaniczna, autonomiczna aktywność poznawcza i asocjacyjna, stanowiąca odpowiednik procesów

myślenia.

Taki świadomy komputer miałby utworzony w znaczącej przez konotację sieci swoich elementów obraz świata zewnętrznego, stale uzupełniany przez nowo napływające z tego świata sygnały (stąd: wrażenia, pojęcia, wyobrażenia i inne, dyskutowane wcześniej, subiektywne zjawiska mentalne). Rozwój i funkcjonowanie takiego odwzorowania świata nadzorowałby pewien ośrodek poznawczy. Jego nakierowanie nie tylko na świat zewnętrzny, ale także na samego siebie spowodowałoby wyłonienie się podmiotu, poczucia własnego „ja” i całej subiektywnej sfery zjawisk psychicznych. Dopiero taki samoświadomy podmiot byłby zdolny do stawiania sobie własnych celów i problemów, a więc posiadałby prawdziwą inteligencję.

Nie mogę uwolnić się od wrażenia, iż do powstania samoświadomości i osobowości niezbędne jest także uwikłanie całego systemu w wewnętrzne sprzeczności. Mam na myśli nie tylko czysto poznawczo-logiczne niespójności, ale także konflikt między popędem poznawczym i rozlicznymi innymi popędami (samozachowawczym, przedłużenia gatunku) oraz między osiami przystosowanie-nieprzystosowanie i przyjemność-przykrość. U człowieka owo nagromadzenie sprzeczności i konfliktów stanowi wynik trwającej przez setki milionów lat ewolucji mózgu, gdzie nowe części dodawane były do starych bez żadnego z góry założonego planu, oportunistycznie i niechlujnie wręcz z punktu widzenia perfekcyjnych rozwiązań konstrukcyjnych. Zachodzi jednak pytanie, czy, gdyby zastosowane zostały rozwiązania technicznie optymalne, powstałaby rasa ludzka, w której indywidua obdarzone są osobowością i podmiotowością — wszak irracjonalność oraz nieustanne miotanie się pomiędzy różnymi motywacjami stanowi immanentną wręcz cechę naszego umysłu. Obawiam się więc, że stworzenie komputera obdarzonego psychiką i osobowością wymagałoby „zaimplantowania” mu gromadzonego przez wieki systemu sprzeczności i wewnętrznych konfliktów.

Podsumowując, pragnę stwierdzić, że za kluczowe dla powstania sztucznej inteligencji (a właściwie: sztucznej psychiki) uważam trzy cechy. Chodzi tu o relacyjną, znaczącą przez konotację sieć elementów przetwarzających informację (zawierającą pewną instrumentalną reprezentację zewnętrznego świata), motywacyjny ośrodek popędowy „sprzężony” z układem nagrody/kary (nadający całości intencjonalność), a także pewien samo-nakierowany na siebie system poznawczy (leżący u podłoża samo-świadomości). Oczywiście jest, że żadne z istniejących obecnie urządzeń elektronicznych nie spełnia powyższych warunków.

\*

*Tekst ten to nieco zmodyfikowany Rozdział 7 mojej książki "Od neuronu do (samo)świadomości", Prószyński i s-ka, Warszawa 2005. Angielskie tłumaczenie tej książki ukazało się jako "From neurons to self-consciousness: How the brain generates the mind", Prometheus Books, New York (Amherst) 2011.*

Zobacz także te strony:

[Popędy, emocje, wolna wola](#)

[Powstanie i istota \(samo\)świadomości](#)

---

Przypisy:

[ 1 ] Choć niekoniecznie przy możliwościach technologicznych dostępnych obecnie lub w przewidywalnej przyszłości.

[ 2 ] Rozdział 6 został zamieszczony na łamach *Racjonalisty* jako artykuł "[Powstanie i istota \(samo\)świadomości](#)".

[ 3 ] Moim zdaniem tak nie jest, bo dla każdego *skończonego* zestawu pytań można przygotować zestaw zupełnie mechanicznych odpowiedzi. Dotyczy to także zestawów *sekwencji* powiązanych ze sobą pytań i odpowiedzi. Dlatego test Turinga byłby rozstrzygający jedynie dla *nieskończonego* zestawu pytań, co jest oczywiście pozbawione jakiegokolwiek praktycznego sensu.

#### **Bernard Korzeniewski**

Biolog - biofizyk, profesor, pracownik naukowy Uniwersytetu Jagiellońskiego (Wydział Biochemii, Biofizyki i Biotechnologii). Zajmuje się biologią teoretyczną - m.in. komputerowym modelowaniem oddychania w mitochondriach. Twórca



cybernetycznej definicji życia, łączącej paradygmaty biologii, cybernetyki i teorii informacji. Interesuje się także genezą i istotą świadomości oraz samoświadomości. Jest laureatem Nagrody Prezesa Rady Ministrów za habilitację oraz stypendystą Fundacji na Rzecz Nauki Polskiej. Jako "visiting professor" gościł na uniwersytetach w Cambridge, Bordeaux, Kyoto, Halle. Autor książek: "Absolut - odniesienie urojone" (Kraków 1994); "Metabolizm" (Rzeszów 1995); "Powstanie i ewolucja życia" (Rzeszów 1996); "Trzy ewolucje: Wszechświata, życia, świadomości" (Kraków 1998); "Od neuronu do (samo)świadomości" (Warszawa 2005), [From neurons to self-consciousness: How the brain generates the mind](#) (Prometheus Books, New York, 2011).

[Strona www autora](#)

[Pokaż inne teksty autora](#)

(Publikacja: 24-03-2011)

[Oryginał.](http://www.racjonalista.pl/kk.php/s,1121) (<http://www.racjonalista.pl/kk.php/s,1121>)

Contents Copyright © 2000-2011 Mariusz Agnosiewicz

Programming Copyright © 2001-2011 Michał Przech

Autorem portalu Racjonalista.pl jest Michał Przech, zwany niżej Autorem.  
Właścicielami portalu są Mariusz Agnosiewicz oraz Autor.

Żadna część niniejszych opracowań nie może być wykorzystywana w celach komercyjnych, bez uprzedniej pisemnej zgody Właściciela, który zastrzega sobie niniejszym wszelkie prawa, przewidziane w przepisach szczególnych, oraz zgodnie z prawem cywilnym i handlowym, w szczególności z tytułu praw autorskich, wynalazczych, znaków towarowych do tego portalu i jakiegokolwiek jego części.

Wszystkie strony tego portalu, wliczając w to strukturę katalogów, skrypty oraz inne programy komputerowe, zostały wytworzone i są administrowane przez Autora. Stanowią one wyłączną własność Właściciela. Właściciel zastrzega sobie prawo do okresowych modyfikacji zawartości tego portalu oraz opisu niniejszych Praw Autorskich bez uprzedniego powiadomienia. Jeżeli nie akceptujesz tej polityki możesz nie odwiedzać tego portalu i nie korzystać z jego zasobów.

Informacje zawarte na tym portalu przeznaczone są do użytku prywatnego osób odwiedzających te strony. Można je pobierać, drukować i przeglądać jedynie w celach informacyjnych, bez czerpania z tego tytułu korzyści finansowych lub pobierania wynagrodzenia w dowolnej formie. Modyfikacja zawartości stron oraz skryptów jest zabroniona. Niniejszym udziela się zgody na swobodne kopiowanie dokumentów portalu Racjonalista.pl tak w formie elektronicznej, jak i drukowanej, w celach innych niż handlowe, z zachowaniem tej informacji.

Plik PDF, który czytasz, może być rozpowszechniany jedynie w formie oryginalnej, w jakiej występuje na portalu. **Plik ten nie może być traktowany jako oficjalna lub oryginalna wersja tekstu, jaki prezentuje.**

Treść tego zapisu stosuje się do wersji zarówno polsko jak i angielskojęzycznych portalu pod domenami Racjonalista.pl, TheRationalist.eu.org oraz Neutrum.eu.org.

Wszelkie pytania prosimy kierować do [redakcja@racjonalista.pl](mailto:redakcja@racjonalista.pl)