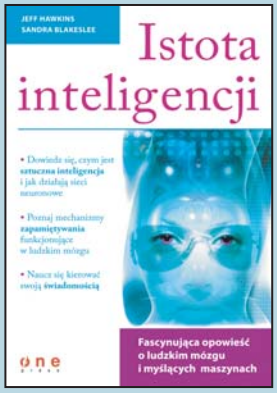
Istota inteligencji



Autorzy: Jeff Hawkins, Sandra Blakeslee
Tłumaczenie: Tomasz Walczak
ISBN: 83-246-0027-2
Tytuł oryginału: On Intelligence
Format: A5, stron: 254

Fascynująca opowieść o ludzkim mózgu i myślących maszynach

- Dowiedz się, czym jest sztuczna inteligencja i jak działają sieci neuronowe
- Poznaj mechanizmy zapamiętywania funkcjonujące w ludzkim mózgu
- Naucz się kierować swoją świadomością

Od lat naukowcy próbują odtworzyć działanie ludzkiego mózgu w maszynach. Sztuczna inteligencja, sieci neuronowe, algorytmy automatycznego wnioskowania – to wszystko owoce tych prób. Jednak żadne z tych rozwiązań nie spełniło pokładanych w nim oczekiwań. Dlaczego? Czy błąd tkwi w nieprawidłowym doborze parametrów sieci neuronowych i błędnej implementacji algorytmów, czy może w fakcie, że próbuje się zamodelować inteligencję bez dokładnego poznania jej tajemnic? Czy jeśli dowiemy się, jak działa nasza inteligencja, będziemy w stanie zbudować prawdziwe „myślące maszyny"?

Dzięki lekturze książki „Istota inteligencji" poznasz odpowiedzi na te pytania. Jeff Hawkins – założyciel firm Palm Computing oraz Handspring – przedstawia w niej najnowsze osiągnięcia grup badawczych zajmujących się analizowaniem mechanizmów kierujących naszą inteligencją, udowadnia, że inteligencja to zdolność mózgu do przewidywania przyszłości przez analogię do przeszłych zdarzeń. Dowiesz się, czym jest inteligencja, jak działa mózg i jak ta wiedza pozwoli zbudować inteligentne maszyny, które nie tylko dorównają możliwościom człowieka, ale nawet je przekroczą.

- Algorytmy sztucznej inteligencji
- Działanie sieci neuronowych
- Funkcjonowanie pamięci w mózgu
- Modele ludzkiej inteligencji
- Zadania kory mózgowej
- Powiązania świadomości i zmysłów artystycznych z inteligencją
- Nowe kierunki badań nad inteligencją

Poznaj przyszłość badań nad sztuczną inteligencją

Wydawnictwo Helion
ul. Chopina 6
44-100 Gliwice
tel. (32)230-98-63
e-mail: helion@helion.pl

SPIS TREŚCI

Mniej więcej w tym samym czasie pojawił się nowy i obiecujący sposób myślenia o inteligentnych maszynach. Sieci neuronowe istniały w pewnej postaci już od późnych lat 60., ale były one konkurencją dla badań Sztucznej Inteligencji, zarówno na polu walki o dotacje, jak i o umysły przyznających je osób. Sztuczna Inteligencja, dominujący kierunek tamtych czasów, aktywnie ograniczała badania nad sieciami neuronowymi. Badacze próbujący pracować nad sieciami zwykle trafiali na długie lata na czarną listę osób, którym nie przyznaje się funduszy. Jednak pewna grupa osób wciąż próbowała rozwijać tę dyscyplinę, aż wreszcie w połowie lat 80. zaświeciło dla nich słońce. Trudno wytłumaczyć, dlaczego akurat w tym okresie nastąpił nagły wzrost zainteresowania sieciami neuronowymi, ale na pewno jednym z czynników były ciągłe niepowodzenia na polu Sztucznej Inteligencji. Zaczęto poszukiwać alternatywy dla tych badań i jedną z nich były właśnie sztuczne sieci neuronowe.

Sieci neuronowe okazały się znacznie lepszym pomysłem od Sztucznej Inteligencji, ponieważ ich architektura jest oparta, choć bardzo luźno, na systemie nerwowym. Zamiast programować komputery, badacze sieci neuronowych, zwani także *koneksjonistami*, chcieli dowiedzieć się, jakie zachowanie mogą przejawiać połączone w grupy neurony. Mózgi składają się z neuronów, dlatego mózg jest w rzeczywistości wielką siecią neuronową. To fakt. Koneksjoniści mieli nadzieję, że uda im się odkryć nieuchwytne właściwości inteligencji, studiując współdziałanie neuronów, a także że dzięki replikacji odpowiednich połączeń między populacjami neuronów uda się rozwiązać niektóre problemy nierozwiązywalne dla klasycznej Sztucznej Inteligencji. Sieci neuronowe różnią się od komputera tym, że nie mają procesora i nie przechowują informacji w scentralizowanej pamięci. Wiedza i wspomnienia sieci są rozproszone, podobnie jak w prawdziwych mózgach.

Na pierwszy rzut oka sieci neuronowe idealnie pasowały do moich własnych zainteresowań. Szybko jednak pozbyłem się złudzeń związanych z tym kierunkiem. Do tego czasu zdążyłem dojść do wniosku, że do zrozumienia mózgu niezbędne są trzy elementy. Pierwszym kryterium było uwzględnienie wśród jego funkcji znaczenia upływu czasu. Rzeczywisty mózg przetwarza błyskawicznie

zmieniające się strumienie informacji. W przepływie danych do mózgu i z niego nie ma nic statycznego.

Drugim kryterium było uwzględnienie informacji zwrotnej. Neuroanatomowie od dawna wiedzieli, że w mózgu istnieje mnóstwo połączeń zwrotnych. Na przykład w sieci połączeń między korą nową a strukturą śródmózgowia zwaną wzgórzem połączenia zwrotne (w kierunku wejścia) są niemal dziesięciokrotnie liczniejsze niż zwykłe połączenia. Oznacza to, że na każde włókno przekazujące informacje do kory nowej przypada dziesięć włókien przekazujących informacje z powrotem do narządów przetwarzających dane zmysłowe. Połączenia zwrotne są dominujące także w obrębie samej kory nowej. Nikt nie zna dokładnej funkcji tych połączeń, ale z opublikowanych badań jasno wynika, że istnieją one we wszystkich częściach mózgu. Sądziłem więc, że muszą być istotne.

Trzecie kryterium to uwzględnienie w teorii lub modelu mózgu jego fizycznej architektury. Kora nowa posiada złożoną strukturę. W dalszej części książki opisuję jej hierarchiczną organizację. Każda sieć neuronowa, która nie uwzględnia istnienia takiej struktury, nie może działać tak jak mózg.

Po eksplozji zainteresowania sieciami neuronowymi dość szybko okazało się, że pozwalają one tworzyć jedynie bardzo proste modele, które nie spełniają żadnego z tych kryteriów. Większość sieci neuronowych składała się z małej liczby neuronów w trzech warstwach. Wzorzec (dane wejściowe) jest wprowadzany do pierwszej warstwy. Znajdujące się w niej neurony są połączone z neuronami drugiej warstwy, tak zwanej warstwy ukrytej. Neurony tej warstwy są z kolei połączone z ostatnią, wyjściową warstwą neuronów, która przekazuje wynik. Siła połączeń między neuronami może się zmieniać, co oznacza, że aktywność jednego neuronu może zwiększać aktywność drugiego neuronu i zmniejszać aktywność trzeciego, w zależności od siły połączeń. Zmieniając wagi połączeń, sieć uczy się przekształcać wzorzec wejściowy we wzorzec wyjściowy.

Takie proste sieci neuronowe przetwarzały jedynie wzorce statyczne, nie używały informacji zwrotnej oraz w ogóle nie przypominały struktur mózgu. Najpopularniejszy typ sieci neuronowych,

zwany siecią „ze wsteczną propagacją błędu", uczy się, przekazując wartość błędu z jednostek wyjściowych z powrotem do jednostek wejściowych. Może Ci się wydawać, że jest to rodzaj informacji zwrotnej, ale w rzeczywistości tak nie jest. Wsteczna propagacja błędu ma miejsce jedynie na etapie uczenia sieci. Kiedy sieć działa w zwykłych warunkach, po fazie uczenia informacje są przekazywane tylko w jedną stronę. Nie istnieje przepływ informacji z warstwy wyjściowej do warstwy wejściowej. Modele te nie uwzględniają także upływu czasu. Statyczny wzorzec wejściowy zostaje po prostu przekształcony w statyczny wzorzec wyjściowy. Następnie przedstawiany jest kolejny wzorzec wejściowy. W sieci nie ma zapisu tego, co zdarzyło się nawet małą chwilę wcześniej. Ponadto architektura tych sieci jest banalna w porównaniu ze złożoną i hierarchiczną strukturą prawdziwego mózgu.

Miałem nadzieję, że naukowcy szybko zaczną się zajmować bardziej realistycznymi sieciami, ale tak się nie stało. Ponieważ proste sieci neuronowe potrafiły wykonywać ciekawe operacje, badania na wiele lat zatrzymały się na tym początkowym etapie. Sieci neuronowe okazały się nowym i ciekawym narzędziem, a w chwilę potem tysiące naukowców, inżynierów i studentów zaczęło dostawać dotacje na badania oraz pisać prace doktorskie i książki na ich temat. Zaczęły powstawać firmy używające sieci neuronowych do przewidywania kursu akcji, analizy wniosków kredytowych, weryfikacji podpisów i wykonywania setek innych operacji wymagających klasyfikacji wzorców. Choć cele twórców tego kierunku mogły być bardziej ogólne, obszar badań został zdominowany przez ludzi, którzy nie są zainteresowani sposobami funkcjonowania mózgu lub zrozumieniem tego, czym jest inteligencja.

Media nie rozumiały zbyt dobrze tych różnic. W gazetach, magazynach i programach popularnonaukowych sieci neuronowe przedstawiane były jako „podobne do mózgu" lub działające „według tych samych zasad, co mózg". W przeciwieństwie do klasycznej Sztucznej Inteligencji, gdzie wszystko musi być zaprogramowane, sieci neuronowe uczyły się na przykładach, co wyglądało na bardziej inteligentne zachowanie. Jednym z najbardziej spektakularnych projektów był NetTalk. Ta sieć neuronowa nauczyła się przekształ-

cać ciągi liter na dźwięki mowy. Trenując generowanie dźwięków na podstawie pisanego tekstu, zaczynała ona brzmieć jak komputerowy głos czytający słowa[1]. Łatwo można było sobie wyobrazić, że po niedługim okresie czasu komputery będą mogły rozmawiać z ludźmi. NetTalk została niezasłużenie okrzyknięta jako maszyna ucząca się czytać. To prawda, że jej działanie było bardzo widowiskowe, ale w rzeczywistości wykonywała zadanie zakrawające na banał. Sieć ani nie czytała, ani nie rozumiała tekstu, przez co jej praktyczne zastosowanie było znikome. Dopasowywała jedynie kombinacje liter do uprzednio zdefiniowanych dźwięków.

Pozwól mi przedstawić analogię, która pokazuje, jak dużo brakuje sieciom neuronowym do prawdziwego mózgu. Wyobraź sobie, że zamiast próbować zrozumieć funkcjonowanie mózgu, starasz się poznać działanie komputera. Po latach badań odkrywasz, że wszystko w komputerze zrobione jest z tranzystorów. Istnieją w nim setki milionów tranzystorów i są one połączone w precyzyjny oraz skomplikowany sposób. Nie rozumiesz jednak, jak komputer działa lub dlaczego tranzystory są połączone w taki, a nie inny sposób. Pewnego dnia decydujesz się więc połączyć kilka tranzystorów, aby zobaczyć, co się stanie. I, uwaga — odkrywasz, że już trzy odpowiednio połączone tranzystory działają jak wzmacniacz. Słaby sygnał przekazany z jednej strony staje się silniejszy w momencie wyjścia z drugiej. (Wzmacniacze w odbiornikach radiowych i telewizyjnych są zrobione z tranzystorów działających w taki właśnie sposób). Jest to istotne odkrycie, a w krótkim okresie czasu powstaje gałąź przemysłu zajmująca się produkcją odbiorników radiowych i telewizyjnych oraz innych elektronicznych urządzeń wykorzystujących tranzystory. Wszystko bardzo dobrze, ale dalej nie wiesz, jak działa komputer. Chociaż zarówno wzmacniacz, jak i komputer są zrobione z tranzystorów, nie mają ze sobą prawie nic wspólnego. Podobnie rzeczywiste mózgi i trójwarstwowe sztuczne sieci neuronowe zbudowane są z neuronów, nie łączy ich niemal nic innego.

[1] Chodzi tutaj o syntezę mowy — *przyp. red.*

Latem 1987 roku miało miejsce zdarzenie, które do reszty osłabiło mój i tak mały entuzjazm dla sieci neuronowych. Uczestniczyłem w konferencji, na której jedną z prezentacji prowadzili przedstawiciele firmy Nestor. Firma ta próbowała sprzedać sieć neuronową służącą do rozpoznawania pisma ręcznego. Licencja na ten program kosztowała milion dolarów, co przykuło moją uwagę. Chociaż firma promowała skomplikowane algorytmy sieci neuronowych i reklamowała je jako kolejny wielki przełom, czułem, że problem rozpoznawania pisma ręcznego można rozwiązać w łatwiejszy, bardziej tradycyjny sposób. Wróciłem tej nocy do domu i zacząłem się zastanawiać nad tym problemem. W przeciągu dwóch dni zaprojektowałem narzędzie służące do rozpoznawania pisma ręcznego. Było szybkie, małe i elastyczne. Moje rozwiązanie nie korzystało z sieci neuronowych ani nie działało jak mózg. Konferencja ta wzbudziła we mnie zainteresowanie projektowaniem komputerów z interfejsem w postaci piórka (co po dziesięciu latach doprowadziło do powstania PalmPilot), przekonała mnie również, że sieci neuronowe nie stanowią wielkiego przełomu w porównaniu z tradycyjnymi metodami. Stworzona przeze mnie metoda rozpoznawania pisma ręcznego stała się podstawą systemu wprowadzania danych o nazwie Graffiti, używanego w pierwszych seriach produktów Palm. Jeśli chodzi o firmę Nestor, to wydaje mi się, że zbankrutowała.

Tyle o prostych sieciach neuronowych. Większość rozwiązywanych przez nie zadań można rozwiązać również tradycyjnymi metodami, a szum medialny wokół nich w końcu ucichł. Przynajmniej twórcy sieci neuronowych nie twierdzili, że ich modele są inteligentne. W końcu sieci te były niezwykle proste i potrafiły mniej, niż programy sztucznej inteligencji. Nie chciałbym pozostawić Cię z przekonaniem, że wszystkie sieci neuronowe to proste trzywarstwowe modele. Niektórzy naukowcy kontynuowali prace nad rozwojem innych rodzajów sieci. Dziś pojęcie *sieci neuronowe* oznacza zróżnicowany zestaw modeli, a niektóre z nich są zbliżone do biologicznego oryginału. Jednak prawie żadne z nich nie próbują odzwierciedlić funkcji lub architektury kory nowej.

Moim zdaniem największy problem z sieciami neuronowymi wynika z właściwości, którą dzielą ze Sztuczną Inteligencją. W obu tych dziedzinach uwaga jest niesłusznie skierowana na samo zachowanie. Nieważne, czy zachowanie to nazwać „odpowiedzią", „wzorcem" czy „wyjściem", zarówno badacze Sztucznej Inteligencji, jak i sieci neuronowych zakładają, że inteligencja polega na zachowaniu, jakie program lub sieć generuje w odpowiedzi na dane wejściowe. Najważniejszą cechą programu lub sieci staje się przez to wygenerowanie poprawnych lub oczekiwanych danych wyjściowych. Zgodnie z nurtem zapoczątkowanym przez Alana Turinga — inteligencja równa się zachowanie[2].

Jednak inteligencja to nie tylko kwestia inteligentnego zachowania lub działania. Zachowanie jest przejawem inteligencji, ale nie jego główną właściwością ani podstawową cechą definicyjną. Przekonuje o tym prosty przykład — możesz być inteligentny, leżąc w ciemnym pokoju i myśląc. Ignorowanie tego, co dzieje się *wewnątrz* głowy, i koncentracja uwagi na zachowaniu zewnętrznym jest wielką przeszkodą na drodze do zrozumienia inteligencji i zbudowania inteligentnych maszyn.

Zanim zaproponuję nową definicję inteligencji, chciałbym przedstawić odmienną wersję podejścia koneksjonistycznego, która jest dużo bardziej zbliżona do sposobu działania rzeczywistego mózgu. Problem polega na tym, że niewiele osób zdaje sobie sprawę ze znaczenia tych badań.

Podczas gdy sieci neuronowe pławiły się w świetle reflektorów, mały odłam teoretyków pracował nad budową sieci, których najważniejszą cechą nie miało być zachowanie. Tak zwana pamięć autoasocjacyjna także była zbudowana z prostych „neuronów", połączonych ze sobą i generujących impuls, kiedy ich pobudzenie przekroczy pewien próg. Jednak wśród ich połączeń znajdowało się wiele połączeń zwrotnych. Zamiast przekazywać informacje tylko

[2] Lecz jeśli chodzi o ścisłość, w myśl teorii Turinga zachowanie pojęte wąsko — w ramach umiejętności prowadzenia spójnego dialogu — *przyp. red.*

w jedną stronę, jak w sieciach ze wsteczną propagacją, pamięć autoasocjacyjna przekazywała impulsy z każdego neuronu także z powrotem, co przypomina nieco rozmowę przez telefon z samym sobą. Kiedy do sztucznych neuronów został przekazany wzorzec aktywności, sieć zapamiętywała ten wzorzec. Sieć autoasocjacyjna kojarzyła wzorzec z nim samym, stąd nazwa — *pamięć autoasocjacyjna*.

Wyniki takiego sposobu łączenia mogą na początku wydawać się zaskakujące. Aby otrzymać wzorzec przechowywany w takiej pamięci, musisz najpierw przekazać go do sieci. Wyobraź sobie, że idziesz do sklepu spożywczego i prosisz o kiść bananów. Kiedy sprzedawca pyta Cię, jak chcesz zapłacić, proponujesz mu zapłatę w bananach. Możesz zadać pytanie, jaki pożytek z takiej sieci? Jednak pamięć autoasocjacyjna posiada kilka ważnych właściwości, które można odnaleźć również w prawdziwym mózgu.

Najważniejsza właściwość to fakt, że w momencie, kiedy chcesz uzyskać dostęp do zapisanego wzorca, nie musisz posiadać do niego pełnego dostępu. Wystarczy jego fragment, a nawet jego uszkodzona wersja. Pamięć autoasocjacyjna potrafi wygenerować cały wzorzec w jego oryginalnej postaci, nawet jeśli do odzyskania użyta zostanie jego uszkodzona wersja. Wyobraź sobie, że idziesz do sklepu spożywczego z połówką brązowego banana i dostajesz w zamian cały zielony owoc. Albo idziesz do banku z poszarpanym i nieczytelnym banknotem, a bankier mówi: „wydaje mi się, że to uszkodzony banknot stuzłotowy. Proszę mi go dać, a ja wydam pani nowiutki, właśnie wydrukowany banknot".

Po drugie, w przeciwieństwie do większości innych sieci neuronowych, pamięć autoasocjacyjna może zostać zaprojektowana w taki sposób, aby przechowywać ciągi wzorców lub wzorce czasowe. Jest to możliwe dzięki połączeniom zwrotnym realizowanym z pewnym opóźnieniem. Dzięki temu możesz zaprezentować pamięci autoasocjacyjnej serię wzorców podobną do melodii, a ona ją zapamięta. Po zapamiętaniu jakiejś melodii, na przykład „Wlazł kotek na płotek", wystarczy zanucić kilka pierwszych nut, a pamięć odtworzy całą melodię. Do odtworzenia całego ciągu wystarczy prezentacja jego części. W dalszej części książki opisuję, że ludzie uczą się praktycznie wszystkiego jako serii wzorców. Uważam, że mózg używa w tym celu podobnych obwodów jak pamięć autoasocjacyjna.

Pamięć autoasocjacyjna wskazywała na wagę połączeń zwrotnych oraz zmiany danych wejściowych w zależności od czasu. Jednak zdecydowana większość naukowców zajmujących się Sztuczną Inteligencją, sieciami neuronowymi i kognitywistyką[3] zignorowała te właściwości.

Naukowcy zajmujący się mózgiem też nie wypadli lepiej pod tym względem. Wiedzieli oni o połączeniach zwrotnych, w końcu sami je odkryli, ale nie byli w stanie wypracować teorii (oprócz niejasnego wspominania o „fazach" i „modulacji"), która wyjaśniałaby, dlaczego znajdują się one w mózgu w tak dużej ilości. Także czas odgrywał małą lub żadną rolę w większości modeli funkcjonowania mózgu. Postrzegali oni mózg przez pryzmat tego, co w jakim miejscu się dzieje, a nie kiedy lub jak wzorce impulsów współistnieją ze sobą. Część takiego nastawienia wynika z ograniczeń współczesnych technik eksperymentalnych. Jedną z ulubionych metod lat 90., zwanych też „dekadą mózgu", było obrazowanie funkcjonalne. Polega ono na tworzeniu obrazów aktywności ludzkiego mózgu, jednak nie pozwala wychwycić szybkich zmian. Dlatego naukowcy proszą badanych, aby skoncentrowali się na pojedynczym zadaniu, podobnie jak fotograf prosi klientów o przybranie odpowiedniej pozy. Dzięki tym badaniom istnieje mnóstwo danych o tym, które obszary mózgu odpowiedzialne są za konkretne zadania, jest jednak bardzo mało informacji o rzeczywistym, zależnym od czasu przepływie informacji. Obrazowanie funkcjonalne pozwala dowiedzieć się, które fragmenty mózgu są szczególnie pobudzone przy rozwiązywaniu danego zadania, ale nie jest w stanie uchwycić zmian jego aktywności w krótkich okresach czasu. Takie dane są potrzebne naukowcom, ale istnieje bardzo mało dobrych technik pozwalających na taki pomiar. Dlatego badania głównego nurtu neuropsychologii wciąż narażone są na błąd związany z analizą samych danych wejściowych i wyjściowych. Prezentuje się badanym stałe bodźce i patrzy, jak reagują. Nałożenie tych danych na mapę mózgu pokazuje, że dane przekazywane są najpierw do pierwszorzędowej kory sensorycznej, gdzie trafiają obrazy, dźwięki i bodźce

[3] Również nauki kognitywne, od ang. *Cognitive Science* — nauki o umyśle i poznawaniu — *przyp. red.*

dotykowe. Następnie trafiają one do obszarów związanych z analizą, planowaniem oraz generowaniem reakcji motorycznych, które przekazują instrukcje do mięśni. Najpierw odbierasz bodziec, potem działasz.

Nie chcę przez to powiedzieć, że wszyscy ignorują czas i połączenia zwrotne. Jest to tak rozległa dziedzina badań, że prawie każda idea posiada swoich zwolenników. W ostatnich latach naukowcy zaczęli przywiązywać większą wagę do połączeń zwrotnych, czasu i przewidywania. Jednak popularność Sztucznej Inteligencji i klasycznych sieci neuronowych przez długie lata ujmowała znaczenia innym nurtom.

Nietrudno zrozumieć, dlaczego ludzie, zarówno laicy, jak i eksperci, uważali, że zachowanie powinno definiować inteligencję. Przez kilka ostatnich stuleci ludzie porównywali możliwości mózgu do mechanizmu zegara, potem do pomp i przewodów, następnie do silników parowych, aż w końcu do komputerów[4]. Literatura i filmy z gatunku fantastyki naukowej przepełnione są ideami sztucznej inteligencji, a od trzech praw robotyki w książkach Isaaca Asimova po C3PO z „Gwiezdnych wojen" obraz inteligentnych maszyn *wykonujących czynności* jest nieodłącznie związany z poruszającymi się robotami. Wszystkie maszyny, wytworzone lub wymyślone przez ludzi, mają coś robić. Nie posiadamy maszyn, które myślą, tylko takie, które działają. Nawet kiedy obserwujemy innych ludzi, zwracamy uwagę na ich zachowanie, a nie na ukryte myśli. Dlatego intuicyjnie wydaje się oczywiste, że inteligentne zachowanie powinno być miarą inteligencji.

Jednak analizując historię nauki, można zauważyć, że podobne intuicje często były największą przeszkodą na drodze do poznania prawdy. Prawdy naukowe często trudno jest odkryć nie dlatego, że są bardzo skomplikowane, ale dlatego, że intuicyjne, lecz niepoprawne założenia nie pozwalają dojrzeć prawidłowej odpowiedzi.

---

[4] A jeszcze wcześniej, w starożytnej Grecji, do katapulty; czy telegrafu w epoce jego rozwoju — *przyp. red.*

Astronomowie przed Kopernikiem (1473 – 1543) niesłusznie zakładali, że Ziemia jest stałym centrum wszechświata, ponieważ *czuli*, że się nie porusza i *wyglądało* na to, że znajduje się w centrum. Było intuicyjnie oczywiste, że gwiazdy są częścią wielkiej ruchomej sfery z Ziemią jako centrum. Gdybyś spróbował zasugerować, że Ziemia obraca się jak bąk, jej powierzchnia porusza się z prędkością ponad tysiąca kilometrów na godzinę, a cała planeta płynie przez przestrzeń, nie wspominając nawet o astronomicznych odległościach od gwiazd rzędu trylionów kilometrów, zostałbyś uznany za niepoczytalnego. Okazuje się jednak, że właśnie tak wygląda rzeczywistość. Łatwa do zrozumienia, ale niezgodna z intuicją.

Przed Darwinem (1809 – 1882) wydawało się oczywiste, że gatunki posiadają stałą postać. Krokodyle nie mogą się krzyżować z kolibrami. Gatunki te są różne i nie można ich połączyć. Pomysł, że gatunki ewoluują, był nie tylko niezgodny z naukami Kościoła, ale także ze zdrowym rozsądkiem. Ewolucja oznacza, że posiadasz wspólnego przodka z wszystkimi organizmami żyjącymi na Ziemi, włączając w to robaki i kwiatki w kuchni. Teraz wiemy, że jest to prawdą, ale jest to sprzeczne z intuicją.

Wspomniałem o tych słynnych przykładach, ponieważ uważam, że próba zbudowania inteligentnych maszyn także jest obciążona intuicyjnymi założeniami, które hamują postęp. Kiedy spytasz się sam siebie, co robią inteligentne maszyny, oczywiste jest, że udzielisz odpowiedzi, opisując określone zachowanie. Ludzką inteligencję opisujemy, przedstawiając mowę, pismo i działania, prawda? Tak, ale tylko do pewnego momentu. Inteligencja to coś, co dzieje się w Twoim mózgu. Zachowanie to tylko dodatkowy składnik. Nie jest to intuicyjnie oczywiste, ale dość łatwo to zrozumieć.

Wiosną 1986 roku, kiedy codziennie przesiadywałem przy biurku, czytając artykuły naukowe, opracowując historię inteligencji i przyglądając się ewolucji Sztucznej Inteligencji i sieci neuronowych, poczułem, że pogrążam się w szczegółach. Cały czas zdobywałem nową wiedzę, ale nie potrafiłem zrozumieć, jak działa mózg jako całość ani nawet co robi. Wynikało to z nadmiernej ilości

szczegółów z dziedziny neuropsychologii, która cierpi z tego powodu do dziś. Każdego roku publikowane są tysiące doniesień z badań, ale najczęściej powodują one przyrost ilości szczegółów, a nie ich lepszą organizację. Wciąż brakuje ogólnej teorii, jakiegoś schematu, który wyjaśniałby, co i w jaki sposób robi mózg.

Zacząłem wyobrażać sobie, jak mogłoby wyglądać takie rozwiązanie. Czy będzie niezwykle złożone, ponieważ mózg jest bardzo skomplikowany? Czy do opisu funkcjonowania mózgu potrzebne będzie stustronicowe dzieło naszpikowane wzorami matematycznymi? Czy zanim cokolwiek zrozumiemy, trzeba będzie utworzyć mapę setek lub tysięcy odrębnych obwodów neuronalnych? Wydawało mi się, że musi istnieć inna droga. Historia pokazuje, że najlepsze rozwiązania problemów naukowych są proste i eleganckie. Podczas gdy szczegóły mogą być skomplikowane, a droga do gotowej teorii wyboista, to ostateczny schemat jest zazwyczaj łatwy do zrozumienia.

Bez wytłumaczenia podstaw, które mogłyby kierować poszukiwaniami, neuropsychologowie nie mają punktu oparcia, który pozwoliłby zebrać wszystkie szczegółowe wyniki badań w jeden spójny obraz. Mózg jest niesamowicie złożoną, nieprzebraną i onieśmielającą plątaniną komórek. Na pierwszy rzut oka wygląda jak stadion pełen spaghetti. Czasem opisywany jest też jako „koszmar elektryka". Jednak po bliskiej i starannej analizie można się przekonać, że budowa mózgu nie jest przypadkowa. Istnieje w nim określona organizacja i struktura, jednak złożoność jest zbyt duża, aby móc intuicyjnie określić ogólne działania mózgu, w przeciwieństwie do odłamków rozbitej wazy, które potrafimy złożyć z powrotem. Problem nie polega na braku wystarczającej ilości danych lub ich jakości. Potrzebna jest zmiana perspektywy. Umieszczone w odpowiednim schemacie szczegóły staną się sensowne i będzie można je zinterpretować. Aby zrozumieć, co mam na myśli, zastanów się nad poniższą analogią.

Wyobraź sobie, że pewnego dnia ludzie wyginęli, a na Ziemi wylądowali przedstawiciele zaawansowanej technologicznie obcej cywilizacji, którzy chcą się dowiedzieć, jak żyli pierwotni mieszkańcy planety. Są szczególnie zainteresowani siecią dróg, pokry-

wającą lądy. Do czego służyły te dziwaczne, wymyślne struktury? Rozpoczynają od skatalogowania wszystkiego za pomocą satelitów i naziemnych pomiarów. Są bardzo skrupulatnymi archeologami. Zapisują położenie każdego fragmentu asfaltu, każdego przewróconego i przeżartego przez rdzę znaku drogowego, każdy szczegół, który uda im się odnaleźć. Zauważają, że niektóre sieci dróg różnią się od pozostałych. W niektórych miejscach posiadają wiele zakrętów, są wąskie i wyglądają, jakby były rozmieszczone niemal przypadkowo. Gdzie indziej tworzą zgrabne, regularne siatki, a w jeszcze innych miejscach stają się dużo szersze i ciągną się setkami kilometrów przez pustynię. Kosmici zbierają mnóstwo szczegółów, które jednak nic dla nich nie znaczą. Zbierają więcej i więcej w nadziei, że jakieś nowe dane pozwolą wyjaśnić wszystko. W takim stanie trwają przez długi czas.

W końcu jeden z nich wpada na pomysł: „Eureka! Chyba wiem, o co chodzi… Oni po prostu nie potrafili się teleportować. Musieli podróżować z miejsca na miejsce, prawdopodobnie za pomocą ruchomych platform o skomplikowanej konstrukcji". Dzięki temu prostemu wyjaśnieniu wiele szczegółów nagle staje się zrozumiałych. Małe, kręte sieci dróg pochodzą z dawnych czasów, kiedy środki transportu były powolne. Szerokie długie trasy służyły do wielogodzinnych podróży z dużą prędkością, co jednocześnie wyjaśnia, dlaczego na znakach znalezionych wzdłuż tych dróg znajdują się inne liczby. Naukowcy zaczynają rozróżniać obszary mieszkalne od przemysłowych, domyślają się wzajemnych wpływów między przemysłem a transportem i rozwiązują inne zagadki. Wiele opisanych szczegółów okazuje się nieistotnych, ponieważ wynikały z historii wypadków lub wymagań stawianych przez ukształtowanie terenu. Istnieje więc taka sama ilość danych, jednak nie stanowią już one zagadki.

Możemy być pewni, że podobny przełom pozwoli nam zrozumieć wszystkie szczegóły mózgu.

Niestety, nie wszyscy wierzą, że można zrozumieć zasady funkcjonowania mózgu. Zaskakująco wiele osób, włączając w to niektó-

rych neuropsychologów, uważa, że w jakiś sposób mózg i inteligencja znajdują się poza możliwością poznania. Niektórzy uważają, że nawet jeśli uda się je zrozumieć, nie będzie możliwe zbudowanie maszyn, które działają w taki sam sposób, ponieważ inteligencja jest nieodłącznie związana z ludzkim ciałem, neuronami i prawdopodobnie jakimiś nowymi i niezrozumiałymi prawami fizyki. Kiedy tylko słyszę te argumenty, wyobrażam sobie intelektualistów z dawnych czasów, którzy sprzeciwiali się studiowaniu nieba lub walczyli przeciw sekcjom zwłok, które ujawniały szczegóły funkcjonowania ludzkiego ciała. „Nie zajmujcie się tymi rzeczami, nie doprowadzą one do niczego dobrego. Nawet jeśli zrozumiecie, jak to działa, nie będziecie mogli w żaden sposób wykorzystać tej wiedzy". Podobne argumenty prowadzą do gałęzi filozofii zwanej funkcjonalizmem, która jest jednocześnie naszym ostatnim przystankiem w krótkiej historii myślenia o myśleniu.

Według funkcjonalistów bycie inteligentnym lub posiadanie umysłu wynika jedynie ze stopnia organizacji materii i nie ma nic wspólnego z tym, jaka jest to materia. Umysł istnieje w każdym systemie, w którym między częściami istnieją odpowiednie związki przyczynowo-skutkowe, a części te mogą równie dobrze być neuronami, jak chipami krzemowymi lub czymś jeszcze innym. Oczywiście taki punkt widzenia jest standardowym podejściem każdego potencjalnego twórcy inteligentnych maszyn[5].

Zastanów się — czy gra w szachy byłaby mniej rzeczywista, gdyby do jej rozegrania użyć solniczek zamiast skoczków? Oczywiście, że nie. Solniczka może być funkcjonalnym odpowiednikiem „rzeczywistego" skoczka dzięki temu, jak się porusza na szachownicy i w jaki sposób wpływa na inne figury, dlatego rozgrywana gra nadal będzie partią szachów, a nie jedynie jej symulacją. Inny przykład — czy to zdanie byłoby takie samo, gdybym najpierw usunął wszystkie znaki, a potem wpisał je na nowo? Albo, używając bliższego każdemu człowiekowi przykładu, zastanów się nad

[5] Jaśniej: według funkcjonalizmu to nie struktura jest istotna (czy to biologiczne neurony, czy krzem), a funkcje, jakie wykazuje dany obiekt. Jeśli komputer będzie się zachowywał tak, jak zachowuje się człowiek w obliczu danej złożonej sytuacji (na przykład przechodząc test Turinga), to *funkcjonalnie* przysługuje mu inteligencja — *przyp. red.*

faktem, że w przeciągu kilku lat wymieniane są wszystkie komórki Twojego ciała. Mimo to pozostajesz sobą. Dany atom jest równie dobry jak każdy inny, jeśli spełnia tę samą funkcję w Twojej budowie molekularnej. To samo dotyczy również mózgu. Jeśli szalony naukowiec postanowiłby wymienić każdy z Twoich neuronów na jego funkcjonalny miniaturowy odpowiednik, to mimo to nadal powinieneś czuć się tą samą osobą co przedtem.

Według powyższego toku rozumowania sztuczny system, który używa tej samej funkcjonalnej architektury co inteligentny, żywy mózg, powinien cechować się równą inteligencją. Nie tylko wymyśloną, ale rzeczywistą, prawdziwą inteligencją.

Badacze Sztucznej Inteligencji, koneksjoniści, a także ja, wszyscy jesteśmy funkcjonalistami, ponieważ uważamy, że w mózgu nie ma nic specjalnego lub magicznego, co pozwala mu być inteligentnym. Wszyscy wierzymy, że w jakiś sposób kiedyś możliwe będzie zbudowanie inteligentnej maszyny. Istnieją jednak różne interpretacje funkcjonalizmu. Powiedziałem już, co uważam za główny problem Sztucznej Inteligencji oraz koneksjonizmu — wynika on z koncentracji na samych danych wejściowych i wyjściowych. Warto też jednak wspomnieć, dlaczego do tej pory nie udało się zaprojektować inteligentnej maszyny. O ile badacze Sztucznej Inteligencji obrali drogę, którą osobiście uważam za zbyt śmiałą, koneksjoniści, moim zdaniem, okazali się zbyt nieśmiali.

Badacz Sztucznej Inteligencji zadaje pytanie: „dlaczego my, inżynierowie, mamy się ograniczać do rozwiązań wymyślonych przez ewolucję?". W zasadzie ma rację. Systemy biologiczne, jak mózg lub genom, są postrzegane jako niezwykle nieeleganckie. Popularną metaforą jest maszyna Rube'a Goldberga, nazwana tak na cześć twórcy komiksów z czasów wielkiego kryzysu, który rysował niesamowicie skomplikowane urządzenia wykonujące najprostsze zadania. Projektanci oprogramowania posiadają podobne pojęcie, *gniot*, oznaczające programy napisane bez odpowiedniego zastanowienia, obarczone nadmierną złożonością, często do tego stopnia, że stają się niezrozumiałe nawet dla ich autorów. Badacze Sztucznej Inteligencji obawiają się, że mózg jest podobnie skomplikowanym gniotem, duszącym się pod naporem kilkuset milionów lat ewolucyj-

nego dziedzictwa. Jeśli tak jest, dlaczego nie porzucić tego wszystkiego i nie zabrać się za projektowanie inteligencji od początku?

Wielu filozofów i psychologów kognitywnych sympatyzuje z tym podejściem. Uwielbiają oni metaforę umysłu jako programu uruchomionego w mózgu, organicznym odpowiedniku sprzętu komputerowego. W komputerach sprzęt i oprogramowanie to dwie zupełnie różne rzeczy. Ten sam program można uruchomić na dowolnej uniwersalnej maszynie Turinga. Możesz włączyć WordPerfect na komputerze klasy PC, Macintosh, a także na superkomputerze Cray, chociaż wszystkie trzy posiadają odmienną konfigurację sprzętową. Sprzęt nie jest też istotny, kiedy próbujesz nauczyć się obsługi programu. Analogicznie można powiedzieć, że mózg nie może nas niczego nauczyć o umyśle.

Obrońcy Sztucznej Inteligencji lubią podkreślać historyczne przykłady, w których rozwiązania proponowane przez inżynierów są zupełnie odmienne od ich naturalnych wersji. Na przykład, jak udało się zbudować samolot? Naśladując machanie skrzydeł ptaków i owadów? Nie, inżynierowie wykorzystali zamocowane na stałe skrzydła oraz śmigła, a później silniki odrzutowe. Wprawdzie rozwiązanie to jest inne niż naturalne, ale działa, i to o wiele lepiej niż machanie skrzydłami.

Na tej samej zasadzie pojazdy, które mogą prześcignąć geparda, nie posiadają czterech nóg, ale koła. Jest to świetny wynalazek do poruszania się po płaskim terenie i nie zmienia tego fakt, że natura nigdy nie wymyśliła takiego rozwiązania. Niektórzy filozofowie umysłu szybko podchwycili metaforę „poznawczego koła", czyli możliwość równie dobrego rozwiązania niektórych problemów w zupełnie inny sposób, niż ma to miejsce w mózgu. Mówiąc inaczej, program, który w danym zadaniu generuje zachowanie przypominające (lub przewyższające) możliwości człowieka, w pewnym ograniczonym, ale użytecznym zakresie jest równie dobry jak mózg.

Wierzę, że interpretacja funkcjonalizmu według zasady „cel uświęca środki" nie doprowadzi badaczy Sztucznej Inteligencji do celu. Jak wykazał Searle, równoważne zachowanie nie jest wystarczającym kryterium. Ponieważ inteligencja jest właściwością mózgu, trzeba do niego zajrzeć, aby się przekonać, czym jest ona naprawdę.

Analizując mózg, a w szczególności korę nową, trzeba uważać na szczegóły, które mogą się okazać jedynie przypadkowymi pozostałościami po ewolucji. Niewątpliwie wśród istotnych cech może znajdować się wiele procesów zbudowanych w stylu Rube'a Goldberga. Ale jak się wkrótce przekonasz, u podłoża obwodów neuronowych znajduje się eleganckie rozwiązanie o mocy przekraczającej możliwości najlepszych komputerów.

Koneksjoniści intuicyjnie czuli, że mózg nie jest komputerem, a jego sekret leży we współdziałaniu połączonych grup neuronów. Był to dobry początek, ale dziedzina ta po początkowych sukcesach praktycznie przestała się rozwijać. Chociaż tysiące osób pracowało nad trójwarstwowymi sieciami neuronowymi, badania nad sieciami przypominającymi naturalne wciąż są bardzo rzadkie.

Od ponad pięćdziesięciu lat ludzie starają się użyć swej wiedzy do zaprogramowania inteligentnych komputerów. W tym czasie udało się utworzyć edytory tekstu, bazy danych, gry komputerowe, Internet, telefony komórkowe i realistyczne animacje dinozaurów, jednak wciąż nie ma śladu inteligentnych maszyn. Aby odnieść sukces, musimy skorzystać z naturalnego nośnika inteligencji, jakim jest kora nowa. Musimy wydobyć inteligencję z mózgu. Inna droga po prostu nie istnieje.