Filozofia Nauki
Rok VIII, 2000, Nr 2(30)

Adam Drozdek

Sieci neuronowe i metodologia nauk

1

Jedną z centralnych kategorii metodologicznych jest pojęcie *teorii*. Jednakże samo określenie, czym jest teoria, rodzi wiele problemów. Czy musi to być system dedukcyjny? Co począć z teoriami w naukach społecznych i humanistyce? Jak się mają synchroniczne i diachroniczne związki pomiędzy teoriami? Kiedy należy teorię uznać za godną przyjęcia? Czy wystarczy poddać ją tylko konfirmacji, czy konieczny jest również proces falsyfikacji? Te i wiele innych pytań doczekało się wielu rozwiązań, które łączy przynajmniej jeden element, mianowicie rozumienie teorii jako zespołu zdań, mniej lub bardziej powiązanych. Dzięki temu zastosować można do nich reguły wnioskowania i dokonywać wyjaśniania, przewidywania oraz badać korespondencje między teoriami. Zasadniczy krok polega więc na przyjęciu, że „przypominające język struktury stanowią podstawową i najważniejszą postać reprezentacji" wiedzy i że „poznanie polega na manipulowaniu tymi reprezentacjami przy użyciu reguł".[1] Wielość proponowanych rozwiązań metodologicznych i trapiące je problemy wskazywać mają, że rozumienie teorii jako zbioru symboli (zdań) jest nieodpowiednie i że teorię należy zdefiniować odmiennie, jako twór o charakterze nie-symbolicznym. Ogromne nadzieje budzi więc rozpowszechniający się paradygmat teorii sztucznej inteligencji, mianowicie konekcjonizm.

---

[1] Paul M. Churchland, *A neurocomputational perspective*, Cambridge: MIT Press 1989, s. 154.

2

Ogólnie rzecz biorąc, konekcjonizm jest strategią badawczą polegającą na wyjaśnianiu zjawisk psychicznych, przede wszystkim zjawiska uczenia się, przy pomocy połączeń powstających między bodźcami a reakcjami w centralnym układzie nerwowym. Konekcjonistyczną teorią jest behawioryzm, który uznawał fakt istnienia tego rodzaju połączeń, lecz nie był zainteresowany ich badaniem, tj. traktował procesy mózgowe jako czarną skrzynkę. Współczesną formą konekcjonizmu jest postulat badania sieci neuronowych. Tak więc czarna skrzynka zostaje otwarta, by można było obserwować procesy w niej zachodzące. Neurologia dokonuje tych badań bezpośrednio, konekcjonizm w teorii sztucznej inteligencji dokonuje tego pośrednio — modelując procesy mózgowe w sieciach neuronowych.

Sieć neuronowa jest zespołem neuronów (węzłów) połączonych między sobą i komunikujących się poprzez połączenia synaptyczne za pośrednictwem przesyłania liczb, których wartość zależy od sygnałów otrzymanych od innych neuronów i od mocy połączeń, tj. wag. Sygnał wysyłany przez dany neuron jest wartością funkcji aktywacji f określonej na sumie ważonej wejść:

$$wyjście = f(waga_1\ wejście_1 + \ldots + waga_n\ wejście_n)$$

Pierwsze sieci neuronowe badane były już w latach czterdziestych. Lata pięćdziesiąte — to początek teorii sztucznej inteligencji, ale bez konekcjonizmu, gdyż pierwsze sieci nie były w stanie poradzić sobie z prostymi nawet zadaniami. Mniej więcej piętnaście lat temu sieci neuronowe weszły ponownie na scenę, gdy się okazało, że teoria sztucznej inteligencji nie rozwija się w obiecywanym początkowo tempie i kierunku, i gdy przezwyciężone zostały pewne problemy nękające pierwsze sieci neuronowe (przede wszystkim przez wprowadzenie sieci wielowarstwowych i użycie sprzężenia zwrotnego przy uczeniu sieci).

Konekcjoniści pragną używać w celu kontrolowania zachowania sieci zamiast reguł wyłącznie połączeń między neuronami. Informacja w sieci reprezentowana jest przez wagi. Jest to reprezentacja rozproszona, tak więc np. liczba, pojęcie lub fraza zakodowane są jako zbiór wag. Sieć zostaje nasycana informacją w procesie uczenia. Istnieje wiele algorytmów uczenia sieci: najpowszechniej używany jest algorytm propagacji wstecznej. Proces uczenia zaczyna się od losowego ustalenia wag połączeń między neuronami; następnie w każdym cyklu procesu uczenia wartości na wyjściu sieci porównane są z wartościami pożądanymi, po czym wagi są odpowiednio modyfikowane, by zmniejszyć różnicę (błąd) między faktycznym i oczekiwanym wyjściem.

Teoria sztucznej inteligencji trapiona jest bardzo poważnymi trudnościami przy budowie systemów umożliwiających np. rozpoznawanie obrazów. Ze względu na nieostrość problemu, algorytmiczne podejście dawało w tym wypadku ograniczone rezultaty. Uczenie sieci neuronowej natomiast nie wymaga określenia algorytmu: wystarczy tu tylko porównywanie wyjścia sieci z oczekiwanym wzorcem. Tysiące a nawet dziesiątki tysięcy cykli w końcu prowadzą do ustabilizowania się wag w sieci tak,

że jest ona w stanie dać pożądane rezultaty. Z tego powodu konekcjonizm jest atrakcyjnym podejściem przy rozwiązaniu problemów, które trudno było rozwiązać w sposób tradycyjny. Zdaniem Churchlanda jest to jedyne rozsądne podejście również w dyskutowaniu problemów metodologii nauk.

3

Otóż paradygmat konekcjonistyczny pozwala na określenie teorii jako „specyficznego punktu w przestrzeni wag synaptycznych" danej sieci, którą jest podmiot poznający. Teoria jest „konfiguracją wag konekcyjnych dzielących aktywacyjną przestrzeń wektorową" na „użyteczne kategorie".[2] Zdefiniowanie teorii jako punktu w przestrzeni wielowymiarowej nie rozwiązuje jednak tradycyjnych problemów epistemologicznych, wciąż bowiem pozostaje otwarty problem interpretacji owego punktu będącego wektorem liczb stanowiących współrzędne tego punktu, gdyż jeśli konekcjonistyczne rozumienie teorii jest właściwe, to na świat trzeba patrzeć jako na szeregi liczb. Tymczasem świat widzimy jako zespół obrazów. Konekcjonistyczne rozumienie teorii musi powiązać przetwarzane przez sieć wektory liczbowe z ich pozaliczbową interpretacją.

Wiadomo ponadto, że różne sesje treningowe dają odmienne konfiguracje wag dla tego samego zbioru wektorów treningowych. Czy mamy wtedy do czynienia z różnymi teoriami? Jeśli zaś nie, jakie jest kryterium podobieństwa różnych teorii, tj. różnych specyficznych punktów? Interpretacja, przy której byłoby tyle teorii dotyczących tej samej sfery rzeczywistości co sesji treningowych, byłaby mało elegancka. Nie trzeba dodawać, że dla różnych struktur sieci teorie będą również odmienne — choć odnosiłyby się do tego samego zbioru wektorów. Aby poradzić sobie z problemem takiej proliferacji teorii, należałoby ustalić kryteria podobieństwa między różnymi wektorami wag. W takim wypadku, wektor byłby po prostu odpowiednikiem reprezentacji rzeczywistości i trudno byłoby przypuścić, że badanie podobieństwa takich wektorów mogłoby przebiegać bez odwołania się do «opisywanej» przez nie rzeczywistości, tj. bez analizy semantycznej. Konekcjonistyczny paradygmat nie byłby wówczas przewrotem w metodologii, gdyż musiałby podjąć trudności tradycyjnego paradygmatu.

4

Sieć neuronowa jest cząstkowym tylko modelem poznania:[3] mianowicie modelem co najwyżej w odniesieniu do mechanizmu postrzegania. W odniesieniu zresztą do

---

[2] *Ib.*, s. 177.

[3] Istnieją poważne wątpliwości, czy w ogóle sieć neuronową można uważać za adekwatny model mózgu. Zob. np.: Jacob T. Schwartz, „The new connectionism: developing relationships between

postrzegania jest to również model cząstkowy. Gdy twarz znajomego odbija się na siatkówce oka, to jesteśmy świadomi faktu, że widzimy twarz znajomego, a nie wektor liczbowy reprezentujący tę twarz. Gdy rozpoznajemy owego znajomego, to rozpoznajemy go jako osobę, a nie jako reprezentujący go wektor liczbowy. Sieć neuronowa operuje na liczbach reprezentujących obraz rozbity na poszczególne fragmenty (piksele). Czy rzeczywiście jednak sieć ta *widzi* obraz? Sieć daje na wyjściu liczby reprezentujące daną osobę. Czy sieć ta jednak *wie*, że właśnie rozpoznała twarz danej osoby? Sieć neuronowa jest w stanie odróżnić twarz jednej osoby od innej; nie staje się jednak tym samym podmiotem obdarzonym władzą poznawczą. Posiada pojęcie twarzy w ogóle i pojęcie twarzy danej osoby w tym samym stopniu, co magnes posiada pojęcie metaliczności i żelazistości materiałów.[4] Poznanie nie ogranicza się bowiem do przetwarzania liczb — czy to na podobieństwo tradycyjnego komputera, czy też na podobieństwo sieci neuronowej. Przetwarzanie liczb jest co najwyżej możliwą implementacją owego poznania, jego substratem, jego materialnym podłożem. W rzeczywistych neuronach mamy do czynienia nie z liczbami reprezentującymi wagi synaptyczne, lecz z procesami chemicznymi i elektrycznymi, przepływem neurotransmiterów itd.; trudno byłoby powiedzieć, że poznanie jest identyczne z odpowiednimi procesami elektrochemicznymi.

Uczenie sieci prowadzi do wygenerowania zespołu reprezentacji na poziomie neuronów w warstwie. Zdaniem Churchlanda w trakcie procesu uczenia sieci, „system teoretyzuje na poziomie jednostek ukrytych, badając przestrzeń możliwych aktywacji wektorów w nadziei znalezienia pewnego podziału lub zespołu podziałów, który potem może wykorzystać warstwa wyjściowa".[5] John McCarthy wyrażał się swego czasu, że termostat posiada przekonania.[6] Podobny charakter ma twierdzenie, że sieć neuronowa ma nadzieję, że poszukuje rozwiązań i że teoretyzuje.

5

Argumentem na rzecz konekcjonizmu ma być fakt, że sieć neuronowa nie wymaga programowania. Jednakże chociaż sieci neuronowych nie programujemy bezpośrednio, to jednak nie zdajemy się na zupełną przypadkowość ich działania. Istnieją

---

neuroscience and artificial intelligence", w: S.R. Graubard (red.), *The artificial intelligence debate*, Cambridge: MIT Press 1988, s. 123–141.

[4] Problem różnicy między zdolnością rozdzielczą i poznaniem przeprowadza Wilfred Sellars w pracy „Empiricism and the philosophy of mind", w: H. Feigl, M. Scriven (red.), *The foundations of science and the concepts of psychology and psychoanalysis*, Minneapolis: University of Minnesota Press 1956, s. 253—329.

[5] Churchland, *op. cit.*, s. 179.

[6] John McCarthy, „Ascribing mental qualities to machines", w: M. Ringle (red.), *Philosophical perspectives in artificial intelligence*, Atlantic Highlands: Humanities Press 1979, s. 161—195, por. na ten temat mój artykuł „Programmabilism: A new reductionism", *Epistemologia* 13 (1990), s. 189—204.

wszak dziesiątki reguł uczenia się, działających na różnych sieciach o różnych strukturach. Różne reguły i różnie sieci stosowane są do różnych zadań, a decyzja co do tego, którą sieć zastosować do danego zadania, nie jest już przypadkowa i wymaga zaprogramowania.[7]

Działanie sieci zależy nie tylko od jej struktury, lecz i od rodzaju wejść i wyjść. Wejścia z kolei zanim staną się wejściami danej sieci (wyjąwszy wejścia będące wynikiem sprzężenia zwrotnego) zostają przetworzone przez inny system; wyjścia są interpretowane również na zewnątrz sieci. Systemy zewnętrzne mogą być również sieciami, lecz wejścia owych sieci zewnętrznych przetworzone zostają przez jeszcze inne systemy zewnętrzne itd. Ostatecznie dochodzimy do systemu, w którym w sposób nieprzypadkowy, zaplanowany, przetworzone zostają surowe dane. Sieć neuronowa nie widzi twarzy, lecz ciąg liczb reprezentujących wartości pikseli (barwa, nasycenie, jasność), na które zostaje rozbity obraz twarzy. Jednakże liczba pikseli (rozdzielczość) i liczba bitów reprezentujących jeden piksel, decyduje o tym, czy rozpoznawanie obrazów będzie uwieńczone sukcesem czy też nie. Przekształcenie obrazu w ciąg liczb i wartości samych tych liczb są niezależne od sieci rozpoznającej obrazy i dokonane zostało przed przystąpieniem do uczenia sieci rozpoznawania obrazów. Owo przekształcenie może być dokonane nie w języku liczb, lecz w języku np. natężeń impulsów elektrycznych reprezentujących piksele. Są to już jednak innego rodzaju wejścia — wymagające innego rodzaju sieci rozpoznającej obrazy.

David Rumelhart i James McClelland zbudowali system tworzący czas przeszły czasowników. Ich zdaniem system ów nie zawiera żadnych reguł formowania czasu przeszłego, a jednak jest w stanie po odpowiednio długim treningu generować właściwą formę czasu przeszłego czasowników regularnych i nieregularnych. W swej obszernej i często cytowanej pracy, zawierającej krytykę tego systemu, Steven Pinker i Alan Prince zwrócili pierwsi uwagę, że choć system Rumelharta—McClellanda nie zawiera reguł użytych *explicite*, to były one wprowadzone przy przygotowywaniu danych wejściowych. „Zasadniczym krokiem zatem jest krok pierwszy: zakodowanie danych."[8]

Ponieważ wszystko jest pod pewnym względem podobne do wszystkiego, kryterium wyróżnionego podobieństwa musi być ustalone przed przystąpieniem do uczenia sieci. Twarz nieznajomego podobna jest do twarzy znajomego, lecz także do księżyca w pełni, do dyni, do czaszki, a również do jeziora (tkanka mięśniowa twarzy zawiera wodę), kreta (ze względu na owłosienie) itd. Uczenie sieci cierpi na to samo ograniczenie, co używanie w latach siedemdziesiątych tzw. mikroświatów przez badaczy sztucznej inteligencji. Zakres możliwości jest z góry ograniczony, tak że system musi osiągnąć przewidywane wyniki. Inteligencja polega na „reagowaniu w ten sam sposób na *odpowiednio* (*relevantly*) podobne wejścia. Jednakże nie sposób uzyskać od-

---

[7] Por. uwagi Jerry'ego Feldmana, które cytuje Jim Jubak w książce *In the image of the brain*, Boston: Little, Brown 1992, s. 180—181.

[8] Steven Pinker i Alan Prince, „On language and connectionism", *Cognition* 28 (1988), s. 73—193.

powiedniego podobieństwa przez odwołanie się do samego podobieństwa".[9] Kryterium odpowiedniości w obu wypadkach zostaje z góry określone, zaplanowane, zaprogramowane z zewnątrz, a więc najtrudniejsze zadanie jest już wykonane przed przystąpieniem do uczenia sieci. Maszyna do tworzenia czasu przeszłego koduje słowa na podstawie schematu podanego przez Wickelgrena w 1969 roku. W schemacie tym każdy fonem scharakteryzowany jest przez dziesięć cech. Ponadto fonem jest ujęty w kontekście fonemu poprzedzającego go i następującego po nim w danym słowie. Ostatecznie każdy fonem reprezentowany jest przez kombinację 33 cech: 30 cech reprezentujących obecności każdej z dziesięciu cech i 3 cechy reprezentujące obecność przerw między słowami. Uwzględnienie wszystkich kombinacji wymagałoby użycia sieci złożonej z $11^3 = 1331$ neuronów, co przekraczało możliwości techniczne i dlatego system zredukowany został do 460 neuronów. Ponieważ system ten niemal w całości polega na kodzie fonologicznym, nie jest w stanie nauczyć się czasu przeszłego w wypadku, gdy przekształcenie czasownika nie jest natury fonologicznej (np. *come* zostaje przekształcony w *came*, lecz *succumb* wcale nie staje się *succame*). System nie jest też w stanie nauczyć się czasu przeszłego homonimicznych czasowników (*ring* — *rang*, *ring* — *ringed*). Tak więc zastosowanie reguł kodowania z jednej strony determinowało odpowiednie działanie systemu, z drugiej zaś ograniczało to działanie.

Pozostaje problem innych pojęć metodologicznych.

6

Według konekcjonistycznego ujęcia teorii, rola teorii sprowadza się do kategoryzacji postrzeżeń zmysłowych, a więc zarówno centralne funkcje teorii, jak i wyjaśnianie i przewidywanie są w tym ujęciu nieobecne. Jest to nieobecność programowa, gdyż wyjaśnianie, jako operacja inferencyjna, jest procedurą wymagającą użycia symboli (języka), a konekcjonistyczna filozofia nauki ma za zadanie właśnie eliminację tradycyjnego podejścia. Nie znaczy to jednak, że pojęcie *wyjaśniania* zostaje na gruncie konekcjonizmu całkowicie wyeliminowane.

Wyjaśnienie zjawiska przypomina proces rozpoznania go, gdyż polega ono na pobudzeniu prototypu (eksplanansu) pozwalającego podmiotowi poznającemu na poradzenie sobie z daną sytuacją (eksplanandum) wymagającą zrozumienia. „Rozumienie wyjaśniające polega na pobudzeniu specyficznego, prototypowego wektora w dobrze wytrenowanej sieci. Polega ono na ujęciu zadania problemowego jako przypadku pewnego ogólnego typu — typu, dla którego podmiot posiada szczegółową i nasyco-

---

[9] William S. Robinson, *Computers, minds, and robots*, Philadelphia: Temple University Press, s. 214. Por. też wypowiedź Pinkera i Price'a: „Jeśli schematy danych odpowiednie do generalizowania nowych form nie są zakodowane w reprezentacji danych, to żadna sieć — a w rzeczy samej, żaden system algorytmiczny — nie będzie w stanie ich odnaleźć"; *op. cit.*, s. 91.

ną informacją reprezentację".[10] Gdy nadjeżdża samochód w naszym kierunku, to postrzegając to — instynktownie uskakujemy. Wyjaśnienie zjawiska polega zaś na odpowiedzeniu na pytanie, dlaczego ów samochód jedzie prosto na nas i podaniu odpowiedzi, że np. hamulce przestały funkcjonować, kierowca jest nietrzeźwy itp. W tradycyjnym ujęciu bardzo wyraźnie odróżnia się proces postrzegania od wyjaśniania. W ujęciu konekcjonistycznym procesy te zlewają się. W wypadku postrzegania i wyjaśniania sieć (podmiot) zaczyna od wektora reprezentującego zjawisko i zmierza do dwóch różnych skutków — reakcji na zjawisko i wyjaśnienia go. Jak to jest możliwe? Otóż to samo wejście prowadzić może w różnych sesjach treningowych do różnych konfiguracji wag; uczenia dokonuje się jednak po to, by wyjście było to samo. Natomiast w wypadku wyjaśniania żąda się, by to samo wejście (wektorowa reprezentacja zjawiska) prowadziło do różnych skutków (reakcji i wyjaśnienia). Jest to możliwe tylko wtedy, gdy reakcji na zjawisko zawsze towarzyszy wyjaśnienie go i odwrotnie, czyli że proces postrzegania jest nieodróżnialny od procesu wyjaśniania. Oczywiste jest jednak, że tak być nie musi. Skąd więc pochodzi różnica? Musi istnieć dodatkowe wejście odpowiadające każdej operacji i uaktywniane przez samą sieć, i właśnie wartość tego wejścia wskazuje, czy daną operację należy dokonać na wejściach reprezentujących dane zjawisko. Jeśli wejście $p = 1$, to zjawisko ma być tylko postrzeżone; jeśli wejście $w = 1$, to zjawisko ma być wyjaśnione. Różnica między postrzeganiem a wyjaśnianiem sprowadzona zostaje w ten sposób do różnicy na owych dodatkowych wejściach. Dlaczego podmiot-sieć uskakuje widząc nadjeżdżający samochód? Właśnie dlatego, że ujmuje „zadanie problemowe jako przypadek pewnego ogólnego typu — typu, dla którego podmiot posiada szczegółową i nasyconą informacją reprezentację". Posiadając tę reprezentację, podmiot przewiduje skutki powstałej sytuacji i dlatego usuwa się z drogi. Ponieważ podmiot jest dobrze wyuczoną siecią, to reakcja jest natychmiastowa. Czym zatem różni się operacja wyjaśniania od postrzegania? Tylko tym, że wyjaśnianie jest nie-instynktowne, a zatem wolniejsze od prostego postrzegania, lub tylko występowaniem refleksji nad reakcją, a nie rozważaniem, czy reakcja była odpowiednia czy nie, tj. czy wyjaśnienie było właściwe czy nie.[11] W każdym razie różnica między wyjaśnianiem a postrzeganiem zaciera się do tego stopnia, że trudno mówić o dwóch różnych operacjach.

7

Konekcjonizm ignoruje też operację wnioskowania: struktura reprezentacji istniejących w sieciach neuronowych po prostu zostaje pominięta. Jednakże, jak zauważają Fodor i Pylyshyn w swym szeroko dyskutowanym artykule, struktura reprezenta-

---

[10] Churchland, *op. cit.*, s. 210.

[11] William Bechtel i Adele Abrahamsen, *Connectionism and the mind*, Oxford: Blackwell 1991, s. 293.

cji jest niezbędna w celu wyjaśnienia trzech elementów poznania: produktywności, systematyczności i spójności inferencyjnej.[12]

Produktywność jest cechą pozwalającą na zakodowanie nawet nieskończonej liczby elementów przy użyciu skończonej liczby symboli. W ten sposób teorię można zdefiniować rekurencyjnie przy użyciu skończonego zbioru przesłanek (lub schematów przesłanek) i reguł inferencji pozwalających na wyprowadzenie nieskończonej liczby wniosków. W paradygmacie konekcjonistycznym definicja taka traci po prostu sens.

Systematyczność umożliwia zrozumienie zdań na tej podstawie, że inne zdania są zrozumiałe. Na przykład reguły syntaktyczne pozwalają uznać, że pewne zdania są elementami danego języka na tej podstawie, że inne zdania należą do tego języka. Przy podejściu konekcjonistycznym nie ma wcale powodu, aby uznawać, że ciąg słów „Maria kocha Jana" jest zdaniem poprawnie zbudowanym w języku polskim dlatego, że zdanie „Jan kocha Marię" należy do tego języka. Zdania są tutaj atomami, a nie strukturami utworzonymi zgodnie z regułami syntaktycznymi; dlatego powiązania między zdaniami są zupełnie innej natury, aniżeli przy założeniu, że gramatyka determinuje zbiór elementów języka. Struktura wyrażeń językowych zostaje zagubiona w połączeniach synaptycznych i nie można już zaliczyć «mechanicznie» strony biernej danego zdania do języka dlatego, że należy do niego strona czynna.

Spójność inferencyjna jest mechanizmem prowadzącym do tego, że „logicznie homogeniczne wnioskowania wykonane są przez odpowiednio homogeniczne procesy obliczeniowe."[13] Tak więc, na przykład, reguły: $((p \rightarrow q) \wedge p) \rightarrow q$, $((p \rightarrow (q \wedge r)) \wedge p) \rightarrow (q \wedge r)$, oraz $((p \rightarrow (q \wedge r \wedge s)) \wedge p) \rightarrow (q \wedge r \wedge s)$ — podpadają pod ten sam schemat i wystarczy zachować pierwszą z nich by wygenerować pozostałe przez proste podstawienie zmiennych. W sieci neuronowej reguły te muszą być reprezentowane oddzielnie. Ponadto niektóre z tych reguł reprezentowane być mogą bez pozostałych, choć wszystkie z nich podpadają pod ten sam schemat.

8

W paradygmacie konekcjonistycznym nieużyteczne okazuje się też pojęcie *prawdy*, gdyż „niejasne jest, czy dobry system przekonań musi pozostawać w jednoznacznej relacji do świata pozajęzykowego."[14] Chodzi tu jedynie o tworzenie „użytecznych kategorii" w procesie uczenia sieci neuronowej. Konekcjonizm okazuje się nową formą pragmatyzmu. W znanej krytyce obliczeniowego modelu poznania Searle pokazuje, że komputer pozostaje na poziomie syntaksy: poziom semantyki jest w nim

[12] Jerry A. Fodor i Zenon W. Pylyshyn, „Connectionism and cognitive architecture", *Cognition* 28 (1988), s. 3—71; Jack Copeland, *Artificial intelligence: a philosophical introduction*, Oxford: Blackwell 1993, s. 245—247.

[13] Fodor i Pylyshyn, *op. cit.*, s. 48.

[14] Churchland, *op. cit.*, s. 157.

nieobecny. Zgodnie ze skrajnie konekcjonistycznym podejściem reprezentowanym przez Churchlanda mówi nie ma sensu nie tylko dyskutowanie semantyki, ale i dyskutowanie syntaksy. Pozostaje tylko badanie połączeń między neuronami i ich stosowalności przy osiąganiu określonych zadań. Problem relacji semantycznych przez to jednak nie znika. Otóż uczenie sieci metodą propagacji wstecznej wymaga porównania faktycznych wyjść z wyjściami oczekiwanymi i obliczania możliwego błędu. Zakłada się więc ostatecznie, że istnieje «prawdziwe» wyjście, tj. wyjście, ku któremu sieć ma zdążać: ocena błędu staje się odpowiednikiem operacji semantycznej.[15]

Z konekcjonizmu zostaje usunięty również aspekt normatywny. Jeśli zdać się na obserwację sieci rozpoczynającej proces uczenia się od przypadkowej konfiguracji wag, to jak uzasadnić konieczność stosowania się do norm metodologicznych? Teoria jest konfiguracją wag, ale jakie warunki powinna spełniać, aby była «dobrą teorią»? Trudno np. badać jej niesprzeczność, chyba że zredukuje się niesprzeczność do poziomu behawioryzmu i będzie obserwować wyjścia sieci. Trudno mówić o intersubiektywności teorii, bo skoro nawet ta sama sieć może mieć różne konfiguracje wag dla tych samych wejść, to tym bardziej różne sieci będą miały różne konfiguracje. Jak ustalić, że mamy do czynienia z tą samą teorię? Raz jeszcze obserwowanie wyjść zdaje się być jedynym rozwiązaniem. Warto zauważyć, że ten sam problem powstaje nie tylko w dziedzinie metodologii nauk. Churchland ujmuje także etykę jako rezultat nabywania przez sieć wzorców, które zostają zakodowane w jej wagach. Odpowiedź na pytanie, „czy proces ten równoważny jest uczeniu się rzeczywistej Prawdy Moralnej, czy socjalizacji"[16] brzmi: socjalizacji. Na gruncie konekcjonizmu nie ma więc miejsca na etykę normatywną, a Prawda Moralna staje się tylko pseudoproblemem. Wiadomo oczywiście, że mózg jest niezbędny do dokonywania wyborów moralnych, jednakże kwestia tego, co dana osoba powinna uczynić w sytuacji, gdy jest w stanie dokonywać takich wyborów, jest kwestią normatywną i wiedza dotycząca struktury i działania mózgu nie na wiele się zda przy rozwiązywaniu takiej kwestii. Podobnie mało pomocna jest wiedza dotycząca mechanizmu wzroku przy ocenie sądów natury estetycznej.[17]

9

Sieci neuronowe znakomicie sobie radzą z problemami, które trudno ująć w słowa, np. z podziałem na kategorie, kiedy trudno wyszczególnić kryteria tego podziału (podział mebli na stoły, krzesła itp., lub podział wyrazów twarzy na smutne, wesołe itp.). Łatwiejszy w takich wypadkach zaprogramowania jest trening na zbiorze przy-

---

[15] Zob. też Jay F. Rosenberg, „Connectionism and cognition", *Acta Analitica* 6 (1990), s. 39—41.

[16] Churchland, *op. cit.*, s. 300.

[17] Virginia Held, „Whose agenda? Ethics versus cognitive science", w: L. May, M. Friedman i A. Clark (red.), *Minds and morals*, Cambridge: MIT Press 1996, s. 74—75.

kładów. Nie usprawiedliwia to jednak prognoz, że sieci neuronowe dostarczą odpowiedzi na wszystkie problemy związane z poznaniem.

Przypominają one prognozy wypowiadane przed laty w odniesieniu do tradycyjnych komputerów. W latach sześćdziesiątych Simon i Newell byli pewni, że zbudowanie w pełni inteligentnego komputera jest kwestią dziesięciu lat, a Turing uważał, że komputer zdolny do zdania jego testu będzie zbudowany w roku 2000. W obu wypadkach były to ekstrapolacje oparte na kilku błyskotliwych sukcesach.

Rzecz w tym, że problemy zachodzące w sieciach neuronowych należą do innego poziomu aniżeli poziom ludzkich operacji poznawczych. Podobnie społeczeństwa ludzkie złożone są z jednostek, lecz złudzeniem okazało się, że procesy społeczne dadzą się zredukować do procesów psychicznych, zachodzących w tych jednostkach. Konstruowanie i modyfikowanie teorii naukowych — to działania ludzkie, lecz odwoływanie się przy opisie struktury tych teorii wyłącznie do procesów psychologicznych a nawet neurologicznych — a tym bardziej np. do mechaniki kwantowej[18] — nie wydaje się najlepszą strategią badawczą. Sami konekcjoniści zresztą przyznają, że „jest bardzo niejasne, czy modele konekcjonistyczne oferują właściwy fundament do modelowania działalności poznawczej człowieka. Podejście konekcjonistyczne bardzo trudno powiązać z metodologiami empirycznymi."[19]

Za takim postawieniem sprawy kryje się twierdzenie, że „implementacja bardzo poważnie ogranicza klasę możliwych architektur kognitywnych."[20] Procesy poznawcze przebiegają w mózgu, a zatem są one ograniczone w czasie i przestrzeni; np. niemożliwe jest przetwarzanie zbyt dużej — nie mówiąc już o nieskończonej — ilości informacji. Istotnie, urazy mózgu mają wpływ na procesy poznawcze. Nie są to konstatacje zbyt odkrywcze. Krytyka Kartezjusza dokonana przez La Mettriego w *Człowieku-maszynie* polegała głównie na odwołaniu się do faktu zachodzenia zależności między procesami fizycznymi a umysłowymi (np. że narkotyki powodują halucynację, a alkohol zwalnia procesy myślowe). Psychologowie widząc w myśleniu «wydzielinę» mózgu i redukując psychologię do neurofizjologii, nie wykraczają poza poziom analizy La Mettriego.[21] Praktycznym i bardzo poważnym efektem tego bywa redukcja psychiatrii do farmakologii i wyłączne użycie leków w leczeniu pacjentów chorych psychicznie. Nikt co prawda nie myśli o podawaniu naukowcom tabletek metodologicznych, lecz redukcja metodologii do neurofizjologii zmierza *de facto*

---

[18] Zob. też Douglas R. Hofstadter, „Analogy-making, fluid concepts, and brain mechanisms", w: A. Clark i P.J.R. Millican (red.), *Connectionism, concepts, and folk psychology*, Oxford: Clarendon Press 1996, s. 238—239.

[19] Paul Smolensky, „On the proper treatment of connectionism", *Behavioral and Brain Sciences* 11 (1988), s. 2.

[20] Nick Chater i Mike Oaksford, „Autonomy, implementation and cognitive architecture: a reply to Fodor and Pylyshyn", *Cognition* 34 (1990), s. 94.

[21] Tak więc np. psychologowie wolą mówić nie o umyśle lecz o umyśle-mózgu jako nierozdzielnej całości; por. np.: J. Allan Hobson, *The chemistry of conscious states*, Boston: Little, Brown 1994, s. 7; Morton F. Reiser, *Memory in mind and brain*, New York: Basic Books 1990, s. 1.

w tym kierunku. Warto jednak zwrócić uwagę na fakt, że choć paradygmat konekcjonistyczny ma się rewolucyjnie różnić od tradycyjnego paradygmatu symbolicznego, to, poza bardzo nielicznymi wyjątkami, implementacji sieci neuronowych dokonuje się na tradycyjnych komputerach o architekturze von Neumanna, a więc na komputerach, które wymagają programowania. Skoro tak, to należy brać pod uwagę możliwość, że i maszyny Turinga można będzie zaimplementować na sieciach neuronowych. Wówczas okaże się, że podejście nie-konekcjonistyczne ma konekcjonistyczną implementację, a więc zasadniczy argument dotyczący różnej implementacji tych dwóch paradygmatów trzeba będzie odrzucić.