

Krystian SOWA

NADZIEJA NA ELEKTRONICZNĄ ŚWIADOMOŚĆ

Wstęp

Pytanie o „inteligentne maszyny” nie jest nowe. W ostatnich latach wzrosło zainteresowanie nimi w związku z dynamicznym rozwojem elektroniki, spektakularnymi osiągnięciami w technologii komputerowej i wykorzystaniem sieci neuronowych. Celem niniejszego artykułu jest skrótowne przedstawienie wybranych kierunków rozwoju badań w dziedzinie sztucznej inteligencji (dalej SI) i związanej z tym problematyki¹. Rozpocznę od omówienia w niej niektórych elementów rozwoju współczesnej techniki komputerowej, a następnie przejdę do prezentacji tzw. „silnego stanowiska” w SI i jego krytyki, próbując odpowiedzieć na pytanie, czy sztucznie „wytworzona” inteligencja jest, lub może być tożsama z inteligencją człowieka. Oprócz tzw. „kierunku symbolicznego” w SI, wykorzystującego klasycznie, softwarowo programowane komputery, przedstawię również „kierunek konekcyjnistyczny”, opierający się na modelowanych elektronicznie sieciach neuronowych.

1. Elektroniczne tło

Będąc dzieckiem, otrzymałem pewnego dnia ciekawą zabawkę — małe płaskie pudełko z rzędami metalizowanych otworów od góry, dwoma przewodami zakończonymi wtyczkami i żaróweczką. Na górną powierzchnię nakładało się kartę pytań i odpowiedzi, np. geograficznych. Jedną z wtyczek wkładało się do otworu z wybranym pytaniem, natomiast drugą szukało pra-

*UWAGA: Tekst został zrekonstruowany przy pomocy środków automatycznych; możliwe są więc pewne błędy, których sygnalizacja jest mile widziana (obi@opoka.org). Tekst elektroniczny posiada odrębną numerację stron.

¹Chciałbym podziękować ks. prof. Michałowi Hellerowi z Wydziału Filozofii Papieskiej Akademii Teologicznej w Krakowie oraz dr. hab. Mariuszowi Flasińskiemu z Instytutu Informatyki Uniwersytetu Jagiellońskiego w Krakowie za cenne uwagi dotyczące omawianych zagadnień.

widłowej odpowiedzi. Sukces sygnalizowany był zapaleniem się żaróweczki. Ciekaw byłem wtedy, skąd zabawka „wie”, że Warszawa naprawdę leży nad Wisłą, a Paryż nad Sekwaną. Zajrzałem do środka i zobaczyłem płataninę przewodów. Gra nazywała się „Mózg elektronowy”.

Kilkadziesiąt lat wcześniej, w roku 1947, właśnie „Mózgiem elektronowym” okrzyknięto pierwszy elektroniczny komputer cyfrowy ENIAC zbudowany przy użyciu lamp elektronowych. Wykonywał ok. 100 działań arytmetycznych zmiennoprzecinkowych na sekundę. Następna, druga generacja komputerów wykorzystywała już tranzystory, trzecia — układy scalone, a czwarta (obecna) — układy scalone wielkiej skali integracji. Osiągane moce obliczeniowe używanych obecnie komputerów dochodzą do 10^9 działań na sekundę. Mówi się dziś o programach badawczych zmierzających do skontruowania komputerów piątej generacji, przetwarzających już nie liczby, lecz ogólnie pojętą wiedzę, oraz o szóstej generacji — neurokomputerach².

W celu osiągnięcia jak największej wydajności pracy komputera, postęp w dziedzinie miniaturyzacji i technologii wytwarzania idzie w parze z rozwojem architektury komputerowej. Działanie najprostszego komputera opiera się na modelu zaproponowanym przez Johna von Neumanna. Na cykl pracy takiego komputera, związany z wykonaniem programu będącego zbiorem rozkazów, składa się pobranie przez procesor odpowiednich danych z pamięci lub urządzeń wejścia, przetworzenie ich według zadanego programu i wysłanie rezultatu do pamięci lub na wyjście. W każdej chwili komputer jest zajęty wykonywaniem tylko jednej czynności; w jednym cyklu pracy wykonywany jest tylko jeden rozkaz. Wzrost efektywności pracy można jednak uzyskać przez równoległe łączenie komputerów i taką konstrukcję programu, by jednostka nadrzędna przekazywała, w miarę możliwości, jak największe fragmenty programu do wykonania jednostkom podrzędnym pracującym równoległe. Na podobnej zasadzie opiera się też konstrukcja mikroprocesorów z tzw. potokową realizacją rozkazów, umożliwiającą komputerowi realizację większej ilości rozkazów w jednym cyklu pracy. Jeżeli, w opisywanym już przykładzie, na cykl pracy komputera składało się pobranie danych, ich przetworzenie i odesłanie wyników, to można tak zaprojektować procesor, aby jednocześnie z przetwarzaniem danych z poprzedniego rozkazu mógł on np. pobierać dane potrzebne przy realizacji następnego rozkazu. Przy takiej

²Por. W. Kołos, *O myśleniu komputerów*, [w:] *Nauka — religia — dzieje*, red. A. Janik, Kraków 1992, s. 41.

organizacji pracy możliwe jest wykonanie do kilkunastu rozkazów w jednym cyklu pracy komputera.

Owoce opisanego postępu technologicznego jest gwałtowny wzrost mocy obliczeniowej komputerów. Postęp techniczny zdaje się tu nie mieć, prócz prędkości światła, innych limitów. Szacuje się, że w 2020 r. jeden komputer biurowy będzie tak wydajny, jak dziś wszystkie komputery razem wzięte³. Maszyny cyfrowe od samego początku służyły człowiekowi pomocą w żmudnych obliczeniach arytmetycznych, jednak od pewnego czasu zaczęły go zastępować, począwszy od „inteligentnego” projektowania i optymalizacji różnych zadań, poprzez diagnozowanie czy prognozowanie, a skończywszy na grze w szachy. Systemy te (tzw. „systemy ekspertowe”), oparte są, najogólniej mówiąc, na nagromadzeniu pewnej wiedzy — utworzeniu bazy danych w odpowiednim języku programowania i podaniu reguł, jak z wiedzy tej korzystać. Jako przykład może służyć tu projekt CYC⁴, którego celem jest przekazywanie komputerowi całości wiedzy człowieka w jasnej, algorytmicznej postaci. Jest to próba poradzenia sobie z klasycznym ograniczeniem systemów ekspertowych, jakim jest ograniczony zakres ich „kompetencji”, to jest pytań, które można im stawiać. Czy jednak da się zalgorytmizować „zdrowy rozsądek” człowieka? Naukowcy zaangażowani w projekt CYC wierzą, że tak.

2. Kierunek symboliczny w Sztucznej Inteligencji

W takim kontekście rodzi się pytanie, czy przykładowy system eksperutowy CYC, działający w oparciu o składający się z ciągu symboli program, można nazwać inteligentnym? Podręcznikowe określenie SI mówi o „dziedzinie nauki usiłującej wytłumaczyć i emulować zachowanie inteligentne w kategoriach procesów obliczeniowych”⁵. Sytuacja byłaby prosta, gdyby mówiło się tylko o naśladowaniu inteligentnego działania człowieka, komplikuje się natomiast, jeżeli stawia się pytanie: czy inteligencja ta jest tożsama z inteligencją ludzką? Zwolennicy tzw. „silnego stanowiska” w SI (dominującego w latach 50–tych) udzielają na to pytanie odpowiedzi twierdzącej. Słowo „sztuczny” jest tu używane nie w znaczeniu „imitacja” (np. sztuczny kwiat), ale w znaczeniu „nienaturalny, lecz nie do odróżnienia od naturalnego” (np. sztuczne jezioro)⁶. Dla silnej SI ludzki mózg jest tylko jedną

³Por. D. A. Patterson, *Mikroprocesor za 25 lat*, „Świat Nauki” 11: 1995, s. 42.

⁴Por. D. B. Lenat, *Sztuczna inteligencja*, „Świat Nauki” 11: 1995, s. 55.

⁵R. J. Schalkoff, *Artificial Intelligence*, McGraw-Hill 1990, s. 2.

⁶Por. W. Kolos, dz. cyt., s. 42.

z realizacji inteligencji, która wcale nie musi być związana z organizmem. Myślenie polega tu tylko na przetwarzaniu danych, a więc inteligencja może realizować się poza człowiekiem. Umysł pełni funkcję *softwaru*, do uruchomienia którego wystarczy dowolny, odpowiednio skomplikowany *hardware*, np. komputer. Kryterium inteligencji komputera byłoby natomiast pozytywne przejście tzw. „testu Turinga”⁷. W teście tym jedna osoba zadaje te same pytania człowiekowi oraz komputerowi i otrzymuje odpowiedzi, nie wiedząc jednakże, która skąd pochodzi. Komputer uważany jest za inteligentny, jeżeli osoba egzaminująca na podstawie otrzymywanych odpowiedzi (analizy funkcjonalistycznej) nie potrafi rozróżnić, „kto jest kto”. Choć obecne „inteligentne” systemy wykonują test Turinga w bardzo ograniczonym zakresie, to ciągły rozwój techniki komputerowej daje zwolennikom silnej SI nadzieje na osiągnięcie ideału.

Krytyka „silnego stanowiska” w SI idzie w dwóch kierunkach. Po pierwsze, analizuje się pracę umysłu oraz maszyny cyfrowej, dokonując rozróżnienia pomiędzy semantyczną pracą umysłu oraz syntaktycznym działaniem komputera. Po drugie, wskazuje się na niealgorytmiczność niektórych procesów zachodzących w ludzkim umyśle w odróżnieniu od algorytmicznej pracy maszyny cyfrowej.

Krytykiem reprezentatywnym dla pierwszego kierunku jest John Searle. Analizując własności umysłu, wyróżnia on takie jego cechy, jak świadomość, intencjonalność, subiektywność i przyczynowe oddziaływanie psychiki⁸. Zastanawiając się zwłaszcza nad problemem odtwarzania stanów intencjonalnych⁹ człowieka, Searle dochodzi do wniosku, że

jeżeli myśl jest zawsze myślą o czymś, przeto dany ciąg symboli musi mieć jakieś znaczenie, by stał się myślą. Mówiąc krótko, umysł ma coś więcej niż syntaktykę, ma semantykę. Powód, dla którego komputerowy program nie może być umysłem, jest prosty: komputerowy program ma cechy syntaktyczne, umysły mają coś więcej, niż syntaktykę. Umysły są semantyczne w tym sensie, że poza strukturą formalną mają jeszcze jakieś treści¹⁰.

Ilustracją tej tezy jest podany przez Searle’a przykład, który wszedł do literatury dotyczącej SI pod nazwą „chińskiego pokoju”. Osoba nie znająca

⁷Por. R. Penrose, *Nowy umysł cesarza*, Warszawa 1995, s. 20.

⁸Por. J. Searle, *Umysł, mózg, nauka*, Warszawa 1995, s. 15.

⁹To znaczy takich, w których umysł ludzki skierowuje się ku obiektowi zewnętrznemu.

¹⁰J. Searle, dz. cyt., s. 28.

języka chińskiego znajduje się zamknięta w pokoju wyposażona w książkę czysto formalnych reguł manipulowania znakami tego języka. W odpowiedzi na otrzymywane z zewnątrz znaki, osoba ta (obrazująca działanie komputera), patrząc na książkę reguł (program komputerowy), wysyła na zewnątrz odpowiednie symbole. Z zewnątrz wygląda to tak, jak gdyby osoba rzeczywiście znała język chiński, choć naprawdę nie rozumie znaczenia żadnego z otrzymywanych czy też wysyłanych znaków. Searle wnioskuje więc, że „wyposażenie jakiegoś zbudowanego przez nas artefaktu w program komputerowy nie wystarczy, aby miał on stany umysłowe porównywalne z ludzkimi”¹¹. Na zarzut, że w podanym przykładzie nie osoba, ale cały system: osoba wraz z książką, z programem i z kosztami symboli rozumie język chiński, Searle odpowiada podobnie. Niemożliwe jest, również dla systemu, przejście od syntaktyki do semantyki. Nie wiadomo, jak system mógłby wyobrazić sobie znaczenie. Człowiek rozumie to, co myśli, natomiast komputer operuje tylko zręcznie symbolami. Komputer naśladowujący człowieka nie będzie nigdy człowiekiem. Żółw elektroniczny¹² pozostanie żółwem elektronicznym.

Stany umysłowe są dla Searlea zjawiskiem biologicznym, choć tajemnicze są owe biologiczne moce, które miałby posiadać mózg ludzki a nie posiada komputer. Wydaje się, że tak złożony układ biologiczny, jakim jest ludzki mózg, posiada więc pewien nieuchwytny fizyko–chemicznie wymiar, pewien „naddatek sensu”, dlatego nie wystarczy rozpatrywanie tylko poziomu fizyko–chemicznego, czy symulacji komputerowej.

Drugi kierunek krytyki wiąże się z komputerową reprezentacją wiedzy. „Inteligentne” systemy ekspertowe stanowią zbiór pewnych danych reprezentujących jakiś fragment świata oraz reguł posługiwania się tymi danymi. Z fragmentaryczności posiadanych danych oraz specyfiki działania (do określonego problemu system stosuje odpowiednie reguły i tak uzyskuje rozwiązanie) wynika ograniczoność tych systemów. Przejawia się ona w istnieniu ściśle określonego zakresu poprawnego działania systemu. Np. system ekspertowy, odpowiadający na pytania z danej dziedziny, będzie znakomicie radził sobie ze skomplikowanymi pytaniami specjalistycznymi, natomiast

¹¹Tamże, s. 37.

¹²Tzw. żółw elektroniczny był jednym z pierwszych urządzeń SI zbudowanym na początku lat 50-tych przez W. G. Waltera. Poruszał się po podłodze, korzystając z własnego napędu, dopóki napięcie w baterii nie spadło dostatecznie nisko, po czym rozpoczynał poszukiwanie najbliższego kontaktu. Po naładowaniu baterii ruszał w dalszą drogę.

nie będzie wiedział, co zrobić z prostym pytaniem wymagającym „zdroworozsądkowej” orientacji w świecie.

Patrząc na człowieka i jego sposób poznawania, czy też podejmowania decyzji, spotykamy się z sytuacjami, w których człowiek po prostu „wie” jak postąpić. Decyzję wydaje się podejmować nie w wyniku stosowania jakichkolwiek reguł, ale właśnie „zdroworozsądkowo”, uwzględniając przy tym cały kontekst sytuacji. Reprezentowanie tej ogólnej wiedzy o świecie, która wydaje się być niewerbalizowalna i nieujmowalna w reguły, stanowi problem dla systemu ekspertowego. Orientowanie się w kontekście sytuacji, tak naturalne dla człowieka, stanowi trudność dla komputera (tzw. *frame problem*). Dochodzą do tego kłopoty komputerów z „myśleniem” indukcyjnym, uogólniającym, choć z dedukcją maszyny cyfrowe radzą sobie bardzo dobrze.

Analizując proces poznawania, Searle dochodzi do wniosku, że w przypadku człowieka kierowanie się regułami nie stanowi podstawy jego działania¹³. Wspomiane już myślenie zdroworozsądkowe, czy metaforyczne wymaga czegoś więcej, niż tylko formalnego programu będącego zbiorem reguł. U podstaw działania świadomości powinien być obecny czynnik niealgorytmiczny.

Do podobnych wniosków dochodzi Roger Penrose, patrząc na zagadnienie od strony matematycznej, przytaczając m. in. wersję twierdzenia Gödla¹⁴. Formulowanie sądów przez człowieka jest operacją niemożliwą do opisanego w postaci programu komputerowego. Czasami do rozstrzygnięcia problemu potrzebny jest tzw. wgląd; np. w matematyce czasem „widać”, czy zdanie jest prawdziwe, czy fałszywe, ale nie można wyprowadzić tego zdania z aksjomatów¹⁵. Ponadto Penrose przytacza opisy różnych sytuacji, wydające się potwierdzać (choć nie stanowiące dowodu) niealgorytmiczność myślenia człowieka: doświadczenia „natchnienia” (rozwiązania przychodzącego w najbardziej nieoczekiwanym momencie), czy też niewerbalności myślenia (myślenie tzw. obrazowe w matematyce czy muzyce)¹⁶. Ciekawy wydaje się również problem wolności woli człowieka, wolności „jakoś” przez człowieka doświadczanej, a niemożliwej do pogodzenia z jego fizycznym opisem¹⁷.

¹³Por. J. Searle, dz. cyt., s. 43.

¹⁴„Dowolny, precyzyjny system aksjomatów i reguł wnioskowania na tyle ogólny, aby obejmował proste zdania arytmetyczne, oraz wolny od sprzeczności, musi zawierać zdania, których prawdziwości nie można ani dowieść, ani obalić, posługując się środkami dozwolonymi w ramach tego systemu” (R. Penrose, dz. cyt., s. 123).

¹⁵Por. tamże, s. 82.

¹⁶Tamże, s. 459, 466.

¹⁷J. Searle, dz. cyt., s. 77.

Krytycy „silnego stanowiska” w SI usiłują, najogólniej mówiąc, wskazać na konieczność uwzględnienia przy opisie człowieka wymiaru wymykającego się analizie fizyko–chemicznej czy algorytmicznej, niemożliwego do elektronicznego odtworzenia. Sądzą, że człowiek to coś więcej, niż tylko działający algorytmicznie komputer.

3. Kierunek konekjonistyczny w Sztucznej Inteligencji

Istnieje jednakże kierunek SI, który inteligencję definiuje inaczej, niż jako umiejętność manipulowania symbolami formalnymi według określonych reguł. Za istotę inteligencji uznano tu nie tyle umiejętność rozwiązywania problemów, ale zdolność uczenia się. Spróbowano stworzyć model ludzkiego mózgu, jako sieci wzajemnie połączonych (stąd nazwa kierunku) neuronów — komórek nerwowych zdolnych do wysyłania i odbierania sygnałów. W ludzkim mózgu każdy z neuronów (a jest ich ponad 10^{11}) może być połączony z innymi za pomocą średnio 10^4 łączników — synaps, co daje całkowitą liczbę połączeń rzędu 10^{15} . W takiej sieci przestaje odgrywać rolę struktura jej elementów składowych — poszczególnych neuronów, ważne są natomiast rodzaje i siły powiązań między nimi. Informacja pamiętana jest więc nie w jakimś pojedynczym neuronie, ale w „całej” sieci — w całości powiązań międzyneuronowych. Tam, gdzie można zastosować sieć, zostaje usunięty jeden z najtrudniejszych problemów podejścia symbolicznego, a mianowicie programowanie komputera. Pisanie programu wiąże się bowiem z koniecznością wcześniejszego dokładnego sformalizowania problemu, co w wielu przypadkach jest bardzo trudne, a czasem niemożliwe. W przypadku pracy wieloprocesorowej dochodzi ponadto problem koordynacji działań poszczególnych procesorów. Tymczasem sieć programuje się „sama”, zdobywając wiedzę w wyniku procesu uczenia się (opis w dalszej części pracy). Rozwiązanie przez sieć zadanego problemu nie pojawia się przy tym wielostopniowo, na skutek analizy programu, ale „od razu”, jednokrotnie, przypominając zdroworozsądkowe reakcje człowieka.

Sieć w zamierzeniu ma być elektronicznym modelem ludzkiego mózgu. W mózgu, chociaż żaden pojedynczy neuron nie posiada świadomości, istnieje ona „całościowo”. Czy więc sztuczne skonstruowanie wystarczająco skomplikowanej sieci wystarczy, by pojawiła się inteligencja tożsama z inteligencją człowieka? Próba odpowiedzi na to pytanie zostanie poprzedzona przedstawieniem kilku podstawowych wiadomości o samych sieciach.

Pierwsze prace dotyczące sztucznych sieci neuronowych (dalej SSN) pojawiły się na początku lat 40–tych XX w. Kierunek ten był z początku

krytykowany przez zwolenników „silnego stanowiska” w SI, preferujących podejście „klasyczne”, związane z podawaniem komputerowi *explicite* algorytmu jego pracy. Dopiero w ostatnich latach badania nad sieciami stały się znów popularne, czy wręcz modne.

SSN można najprościej określić jako sieć połączonych ze sobą procesorów (analogowych lub cyfrowych) nazywanych umownie neuronami. Sztuczne „neurony”, których pierwowzorem są komórki nerwowe budujące mózg człowieka, są skrajnie uproszczone, tak by można było je łatwo budować lub też symulować komputerowo. Już tu napotykamy więc ułatwienia poświęcające dokładność odwzorowania poszczególnych neuronów prostocie ich elektronicznej realizacji. Przyłączeniu takich neuronów w sieć obserwujemy dalsze uproszczenia. SSN buduje się najczęściej jako złożone z pewnej liczby warstw neuronów połączonych pomiędzy sobą w sąsiednich warstwach na zasadzie „każdy z każdym”, wierząc, że w toku nauki sieć sama wypracuje sobie właściwą strukturę. Tak jest najłatwiej. Tymczasem tylko niektóre, małe partie mózgu mają strukturę warstwową, natomiast „logiki” międzyneuronowych połączeń reszty nie udało się odkryć i nic nie wskazuje, żeby nastąpiło to w najbliższym czasie.

Typowy neuron SSN będzie więc elementem posiadającym wiele wejść odbierających sygnały od innych neuronów i jedno wyjście połączone z odpowiednimi wejściami innych neuronów. Sygnał wyjściowy takiego neuronu jest liniową lub nieliniową, deterministyczną lub stochastyczną funkcją sumy ważonej jego wejść. Poszczególne wagi odpowiadają mocy połączeń między poszczególnymi neuronami. Połączenia pomiędzy warstwami mogą być proste: sygnał przechodzi wtedy jednokierunkowo od wejścia do wyjścia przez kolejne warstwy, lub na zasadzie sprzężenia zwrotnego: daje to sieci możliwość rozwiązywania bardziej skomplikowanych problemów. Jeżeli zbiór sygnałów wejściowych sieci jest tego samego rodzaju, co zbiór sygnałów wyjściowych, mówimy wtedy o sieciach autoasocjacyjnych, jeżeli różnego rodzaju — o sieciach heteroasocjacyjnych. Ostatni podział wyjaśnimy na przykładzie rozpoznawania obrazów. Sygnałem wejściowym dla przykładowej sieci niech będzie zbiór elementów fotograficznego obrazu twarzy jakiejś osoby, zniekształcony przez zakłócenia, lub też tylko część tego obrazu. W wyniku działania odpowiednio wytrenowanej sieci autoasocjacyjnej możemy otrzymać odpowiednio nasz obraz bez zakłóceń lub też cały obraz odtworzony na podstawie podanego fragmentu. Sieci heteroasocjacyjnej można by użyć natomiast do klasyfikacji obrazu wejściowego, otrzymując w naszym przypadku na wyjściu sygnał decyzji: np. kobieta/mężczyzna.

Uczenie sieci polega na zmianie połączeń międzyneuronowych i może być przeprowadzane „z nauczycielem” lub też „bez nauczyciela”. W pierwszym przypadku przekazuje się sieci informację wejściową i oczekiwany wynik, a ona sama modyfikuje połączenia (wagi) pomiędzy swoimi elementami, by otrzymać pożądaną wynik (a praktycznie minimalizację wyjściowej funkcji błędu, np. według metody najmniejszych kwadratów). Przy wystarczającej złożoności sieci (np. kilka warstw i odpowiednia ilość neuronów w poszczególnych warstwach) i dostatecznej ilości prezentacji uczących¹⁸, można sieć nauczyć nawet złożonego zachowania się, np. rozpoznawania złożonych obrazów, czy podejmowania skomplikowanych decyzji. W przypadku uczenia „bez nauczyciela”, sieć sama wytwarza swoją wewnętrzną reprezentację sygnału wejścia, modyfikując swoje połączenia np. zgodnie z tzw. regułą Hebba¹⁹: połączenie synaptyczne pomiędzy neuronami A i B wzmacnia się, jeżeli pobudzenie neuronu A powoduje pobudzenie neuronu B , osłabia się natomiast, jeżeli tak nie jest. Sieci takie są wykorzystywane np. do detekcji regularności danych (np. obrazów) wejściowych w sytuacji, gdy nie jesteśmy w stanie przewidzieć, jakie dane mogą pojawić się na wejściu. Jest to analogiczne do sytuacji orientowania się człowieka w otoczeniu, które widzi po raz pierwszy. Przykładem pracy „bez nauczyciela” jest też autoasocjacyjna sieć ze sprzężeniami zwrotnymi, tzw. sieć Hopfielda, która wswiliła się w badaniach nad trudnym obliczeniowo „problemem komiwojażera”²⁰.

Kierunek konekjonistyczny w SI jest więc alternatywny w stosunku do kierunku symbolicznego. SSN i budowane na ich podstawie neurokomputery nie wymagają podawania z góry reguł zachowania się, natomiast wiedzę nabywają w procesie uczenia się oraz są zdolne do samouczenia. Teoretycznie „pojemność” sieci jest nieograniczona, praktycznie zależy od jej struktury i sposobu uczenia. Może się zdarzyć, że sieć, nabywając nową umiejętność, „zapomni” tego, czego nauczyła się wcześniej, ale czy ludzki mózg nie ma podobnych własności? Ważną cechą sieci jest ich niezawodność i odporność

¹⁸Nawet 10¹² przy rozpoznawaniu mowy czy rękopisów (por. R. Tadeusiewicz, *Sieci neuronowe*, „Enter” 3: 1995, s. 85).

¹⁹Hebb analizując pracę mózgu spostrzegł, że zachodzą tam procesy wzmacniania połączeń pomiędzy neuronami pobudzonymi do pracy równocześnie.

²⁰Jest to problem optymalizacyjny. Danych jest n punktów — „miast”, pomiędzy którymi należy znaleźć minimalną drogę globalną, odwiedzając każde miasto tylko raz i wracając do punktu startowego. Matematyczna optymalizacja znalezionej funkcji drogi jest zadaniem trudnym, natomiast sieć sama dokonuje minimalizacji funkcji swojej „energii” wewnętrznej (zależnej od stanów wyjść poszczególnych neuronów i wag poszczególnych połączeń), znajdując rozwiązanie problemu (por. R. Tadeusiewicz, *Sieci neuronowe*, Warszawa 1993, s. 94).

na uszkodzenia. Informacja pamiętana jest w „całej” sieci, stąd utrata jednego, czy nawet kilku neuronów nie powoduje zauważalnego pogorszenia jej pracy. Sieć jest także w stanie, oczywiście w ograniczonym zakresie, uogólnić posiadane wiadomości, np. całkiem nowe.

SSN wykorzystywane są do rozpoznawania i klasyfikacji (np. do celów wojskowych: sieć rozpoznaje bezbłędnie rodzaj samolotu na podstawie małych detali obserwowanych z bardzo dużej odległości, czy też sieć rozróżnia szumy charakterystyczne dla napędu poszczególnych statków i okrętów), sterowania i optymalizacji (np. steruje ramieniem robota przy złożonych manipulacjach), prognozowania (np. przewiduje zapotrzebowanie na energię elektryczną, czy stawia prognozy giełdowe). Sieci stosowane są tam, gdzie formalizacja procesu w postaci programu komputerowego (jeżeli jest w ogóle możliwa) prowadziłyby do bardzo złożonych obliczeń, często niemożliwych do wykonania w czasie rzeczywistym²¹. Wtedy użycie sieci jest prostsze, szybsze, mniej zawodne i tańsze.

Mówiąc o przykładach udanych zastosowań sieci, należy jednakże wskazać również na ich ograniczenia. Jednym z nich jest manipulacja na symbolach formalnych. Przetwarzanie informacji symbolicznej jest, jak na razie, praktycznie nierealizowalne przez sieć. Innym problemem są obliczenia numeryczne o bardzo wysokiej dokładności. Niezależnie od tego, jak spektakularne wydają się zastosowania sieci, ich obliczenia mają zawsze charakter przybliżony, choć wystarczający – jak widać — w wielu zastosowaniach. Problem dla sieci stanowią również obliczenia logiczne wieloetapowe, w których potrzebna jest dokumentacja z poszczególnych etapów ciągu rozumowania, sieć natomiast dąży do jednokrokowego rozwiązania problemu. Tak więc mimo licznych zalet SSN nie grozi nam, przynajmniej na razie, odstawienie klasycznie programowanych komputerów do lamusa.

4. Elektroniczna świadomość?

Pozostaje jednak postawione już wcześniej pytanie: czy skonstruowanie SSN o wystarczającym stopniu złożoności wystarczy, by pojawiła się świadomość, inteligencja? Czy sztuczna sieć neuronów jest zdolna do myślenia twórczego, czy tylko do działania na zasadzie bezmyślnych nawyków będących wynikiem tresury? Na praktyczną odpowiedź na te pytania przyjdzie nam jeszcze długo poczekać. Szacuje się, że do symulowania mózgu potrzebny byłby komputer o mocy obliczeniowej 10^{18} operacji na sekundę,

²¹Por. R. Tadeusiewicz, *Tryumf, czy kapitulacja rozumu*, „Znak” 1995 nr 484, s. 66.

natomiast obecnie dysponujemy komputerami o prędkości rzędu 10^9 operacji na sekundę²². Zostają nam więc rozważania filozoficzne i eksperymenty myślowe.

Pytanie o elektroniczną świadomość jest pytaniem o koncepcję człowieka. Maszyna „jaka jest, każdy widzi”, problem zaczyna się natomiast przy opisie człowieka i jego inteligencji. Jeżeli potraktować człowieka tak, jak np. „posiadający przekonania” termostat z przykładu opisanego przez Searlea²³, to pomiędzy myśleniem człowieka, a funkcjonowaniem komputera czy sieci neuronowej nie będzie różnicy, chyba, że w stopniu skomplikowania. Jeżeli jednak przyjąć koncepcję człowieka myślącego semantycznie i niealgorytmicznie, to staniemy przed pytaniem, czy na podstawie obserwacji działania komputera będziemy mogli przyjąć, że ma ono cechy semantyczne?

Przedstawiony przykład „chińskiego pokoju” Searlea, z którego wynika, że budowane przez człowieka artefakty nigdy nie osiągną semantyki, wydaje się przekonujący w przypadku komputerów algorytmicznych, jednak już nie jest tak oczywisty w przypadku sztucznych sieci neuronowych. W SSN, podobnie jak w mózgu człowieka, informacja nie jest zawarta w pojedynczym neuronie, ale w połączeniach międzyneuronowych. Jeżeli świadomości człowieka nie podlega wątpliwości, to czy na podstawie obserwacji działania SSN można będzie kiedyś powiedzieć, że taka sieć posiada semantykę, czy też doświadcza własnej wolności? Nie chodzi przy tym o wolność w ujęciu kompatybilizmu²⁴, który przyznaje człowiekowi zdolność działania zgodnie z własną wolą, utrzymując jednakże zdeterminowanie tejże woli przez okoliczności. Chodzi o wolność, która byłaby zdolnością człowieka do działania zupełnie nieprzewidzianego, nie zdeterminowanego przez sytuację²⁵.

SSN stanowią rozwiązanie alternatywne w stosunku do klasycznego programowania algorytmicznego, tzn. podawania komputerowi przez programistę mechanicznej procedury, według której przetwarzałyby określone dane. Nie znaczy to jednak, że sieć działa zupełnie niealgorytmicznie. Sieć po

²²Mózg ludzki zawiera ponad 10^{11} neuronów, natomiast konstruowane obecnie SSN zawierają zaledwie do kilkuset tysięcy uproszczonych modeli neuronów sztucznie zgrupowanych w kilka warstw.

²³Searle przytacza tu pogląd McCarthyego, twórcy terminu „sztuczna inteligencja”, który uważa, że nawet tak proste urządzenia jak termostaty posiadają przekonania, a mianowicie sądy: „tu jest zbyt gorąco”, „tu jest zbyt zimno”, „temperatura jest w sam raz” (por. J. Searle, dz. cyt., s. 27).

²⁴Por. tamże, s. 80.

²⁵Searle próbuje pogodzić doświadczenie tej wolności przez człowieka z przyrodniczą wizją mechanicznego świata i przyznaje, że nie bardzo mu się to udaje (por. tamże, s. 77).

prostu „programuje się” sama w wyniku sformalizowanego procesu uczenia się. Działanie sieci jest deterministyczne: w określonych warunkach otrzymujemy określony wynik możliwy do przewidzenia, gdyby dysponować wystarczającymi danymi dotyczącymi stanu sieci. Zdolność sieci do uogólniania posiadanej wiedzy i wykorzystywania jej w nowych sytuacjach, którą obecne sieci posiadają w ograniczonym stopniu, jest więc raczej mechaniczną ekstrapolacją, niż twórczością. Przykład „chińskiego pokoju” Searlea zachowuje swoją ważność również w odniesieniu do sieci neuronowych. Z drugiej jednak strony, uczenie człowieka nie odbiega przecież w niektórych przypadkach od uczenia SSN, a jednak człowiek doświadcza siebie jako istoty świadomej, wolnej i twórczej. Natura świadomości człowieka wciąż pozostaje dla niego tajemnicą. Być może trzeba nam — jak sądzi Searle — dopracować się bardziej wyrafinowanej koncepcji przyczynowości niż ta, którą dotychczas posiadamy²⁶. Być może rozwiązanie problemu przyniesie dopiero dalszy, intensywny rozwój fizyki kwantowej.

Teoretycznie łatwo jest zrobić eksperyment myślowy i przyznać sieci ludzką świadomość. Jest to, o tyle proste, że nie jesteśmy specjalnie kompetentni ani w dziedzinie powstawania świadomości u człowieka, ani w dziedzinie modelowania mózgu — i stąd może nieświadomi wielu trudności, z którymi przyjdzie nam się jeszcze spotkać. Entuzjazm zwolenników poglądu o świadomości SSN może przypominać entuzjazm zwolenników silnej SI w latach 50-tych, kiedy wszystko wydawało się możliwe. Łatwe do wypowiedzenia określenie „sieć o wystarczającym stopniu złożoności” nie zawiera jednak konkretnej treści. Przy zaawansowanych próbach modelowania biologicznego mózgu za pomocą układów scalonych możemy natknąć się jeszcze na niejedyn problem w rodzaju „semantycznej bariery” Searlea. Być może, że do myślenia semantycznego nieodzowny jest właśnie ów nieuchwytny fizyko-chemicznie „wymiar”, który Searle nazywa „czynnikiem biologicznym”, związany z życiem i z cielesnością człowieka, a nie tylko wyabstrahowaną sztuczną siatką neuronów.

A jeżeli nawet skonstruujemy *androida*, który funkcjonalistycznie doskonale replikowałby życie ludzkie, to znaczy posiadałby wszystkie umysłowe i zmysłowe zdolności typowe dla ludzi — z punktu widzenia silnej SI „kiedyś” powinno być to możliwe — to czy można będzie traktować go jak człowieka? Ilustracją problemu niech będzie przykład, w którym mamy do czynienia z dwiema butelkami wina: jedna z nich stanowi oryginalny trunek,

²⁶Por. tamże, s. 18.

a druga jego nieodróżnialną chemiczną podróbką²⁷. Co stanowi o tożsamości pierwszej butelki? Czy tylko smak? Na czym polega przyjemność picia wina? Część uroku wina będzie brała się ze świadomości jego pochodzenia²⁸, zaczynając od dojrzewających na słońcu winogron, a kończąc na procesie tradycyjnego przetwarzania. W stosunku do drugiej butelki reakcje mogą być więc dwojakiego rodzaju. Pierwsza: „smak jest ten sam, ignoruję pochodzenie” i druga „nigdy na moim stole”. Norman Lillegard, odnosząc się do powyższego przykładu, dopuszcza możliwość uznania sztucznie skonstruowanego androida za człowieka na podstawie tylko klasyfikacji funkcjonalistycznej, tak, że pochodzenie przestaje mieć znaczenie, wskazuje jednak na konsekwencje takiego kroku:

Nie twierdzą, że to nie może się zdarzyć. Wierzę natomiast, że można powiedzieć, że gdyby tak miało się zdarzyć, to stracone zostałyby coś bardzo ważnego. Stracilibyśmy pod doniosłymi względami sens samych siebie²⁹.

Lillegard wskazuje na to, że na tożsamość człowieka składa się nie tylko jego ciało w fizycznym, czy nawet biologicznym sensie, ale też cały konkretny kontekst jego życia, narodziny, przemijanie, wreszcie śmierć.

Nasze poczucie bólu jest związane z naszym poczuciem moralności i powinności, cierpienia i perspektywy śmierci. Nakaz, by nie zadawać bólu jest wewnętrznie związany z cnotą dobroci. Czym byłaby ta cnota w świecie nieśmiertelnych [androidów]³⁰.

Czy android będący „poza” sytuacją w rozumieniu człowieka, nie starzejąc się, nie przemijając i nie umierając, byłby zdolny np. do odczuwania bólu? Czy grając w szachy mógłby odczuwać przyjemność z wygranej? Czy jego odczucia byłyby odczuciami człowieka?

Pytanie o inteligencję komputera „o wystarczającym stopniu złożoności” pociąga za sobą następnne: o człowieka i jego inteligencję. Refleksja nad człowiekiem w świetle pytania o SI nie powinna mieć przy tym charakteru

²⁷Por. N. Lillegard, *Sztuczna inteligencja a Biblia*, „Znak” 1995 nr 484, s. 52.

²⁸Lillegard przytacza tu spostrzeżenie, że bytom (np. człowiekowi) nie przypisujemy predykatów psychologicznych tylko na bazie ich działania, ale na bazie tego, co o nich wiemy (por. tamże).

²⁹Tamże, s. 54.

³⁰Tamże, s. 53.

„taniej apologii” człowieka powodowanej lękiem przed utratą jego „wyróżnionej” pozycji w świecie, z drugiej zaś strony, nie powinna być redukcjonistyczna. Pytając o inteligencję człowieka, wypada wziąć pod uwagę nie tylko jego zdolność formalnego rozwiązywania problemów, czy też uczenia się, ale także zdolność rozumienia przyswajanych treści. Pozostaje również cała sfera psychiczna czy aksjologiczna. Człowiek ze swoim biologicznym ciałem, osadzeniem w konkretnym życiu wraz z jego radością i cierpieniem, rozumiejący znaczenie słów takich jak przyjaźń i miłość, wydaje się być czymś więcej, niż tylko sztuczną siecią neuronową o wysokim stopniu złożoności, czy supersystemem ekspertowym. Elektroniczna inteligencja byłaby w takim przypadku tylko funkcjonalistycznym naśladowaniem inteligencji człowieka.

Nie znaczy to, że należy żałować wysiłku na badania w dziedzinie SI, czy to spisując ją na straty, czy też lękając się, że może być konkurencją dla *homo sapiens*. Prace nad SI mogą być pasjonującą przygodą pozwalającą człowiekowi przy okazji zrozumieć lepiej samego siebie. Ale nadzieja na elektroniczną świadomość jest przecież nadzieją człowieka. Czy można mówić zarówno o nadziei, jak i o świadomości w świecie krzemowych androidów?