

I. Czy już żyjemy w erze sztucznej inteligencji?

Pół wieku temu powstały pierwsze komputery w postaci elektronicznych maszyn cyfrowych. Był to początek *cywilizacji informatycznej*. Wraz z nim poczęły rodzić się wizje osiągnięcia stanu, w którym inteligencji ludzkiej nie oddzielałaby od maszynowej magiczna linia posiadania świadomości. Legion maszyn dzielących z ludźmi świadomość, a zarazem wyposażonych w zawrotną moc obliczeniową i zespolonych w globalnej sieci stworzyłby szansę podołania gigantycznym problemom, także tym, od których może zależeć przetrwanie naszego gatunku.

Co sądzić o tych wizjach i jak ma się do nich stan obecny? Od tego pytania zaczniemy rozważania nad maszynową czyli sztuczną inteligencją. Jest mu poświęcony obecny rozdział.

1. Jak doszło do pojęcia sztucznej inteligencji?

Gdy zaczęto rozumieć, że dzieje się coś przełomowego, powstał termin *Artificial Intelligence* — Sztuczna Inteligencja, co skraca się jako *AI*, a w polskim *SI*. Dlaczego wyraża on świadomość przełomu?

Sztuczna inteligencja jest to, w gruncie rzeczy, kolosalna moc obliczeniowa komputera, biorąca się z doskonałości sprzętu i programów. Fachowcy mogliby się więc obejść bez tego terminu; czy by on zaistniał czy nie, rozwój komputerów posuwałby się w tym samym kierunku. Pobudza jednak ten termin do refleksji filozoficznej, która powinna iść w parze z postęпами nauki i techniki. A ponadto, co też ważne, przemawia do masowej wyobraźni.

Zwrot „sztuczna inteligencja” ukazuje coś, czego laik nie wydedukuje z obiegowego pojęcia komputera. Mianowicie, że inne niż rachowanie czynności inteligentne można również powierzyć maszynie — pod warunkiem, że uda się je sprowadzić do rachowania. A to w wielu sprawach dobrze się udaje dzięki osiągnięciom logiki matematycznej i sukcesom elektroniki.

Optymizm jest nie tylko w słowie „inteligencja” lecz i w przydawce „sztuczna”. Trzeba ją brać, oczywiście, nie w tym zgryźliwym sensie, który przypisuje sztuczność protezom i falsyfikatom. Oto jakieś wyższe instancje, Bóg czy Przyroda, stworzyły inteligencję naturalną. Ta zaś okazuje się zdolna wytworzyć inteligencję, będącą nie dziełem natury, lecz ludzkiej sztuki. W tym to sensie, dla człowieka zaszczytnym, jest ona sztuczna.

Maszyna inteligentna, lub *inteligencja maszynowa*, to inne sugestywne określenia. Pojawiły się one w 1948 i 1950 w tekstach Alana Turinga. On pierwszy rozważał na poziomie wykonalności technicznej, jak powierzyć maszynie problemy, których rozwiązywanie było dotąd zastrzeżone dla człowieka.

Alan Turing (1912-1954) wslawił się najbardziej publikacją w roku 1937 podającą po raz pierwszy logiczno-matematyczne podstawy komputera (o którego technicznej realizacji jeszcze nie myślano). Ta konstrukcja, zwana *maszyną Turinga*, po dziś dzień stanowi podstawy teoretyczne SI (będzie o tym mowa w rozdziale trzecim).

Turing wslawił się na inny jeszcze sposób, gdy w czasie drugiej wojny światowej dowodził akcją deszyfrowania niemieckich depech wojennych kodowanych przez słynną „Enigmę”, konstruując do tego

celu coraz doskonalsze maszyny dekodujące. Prace te rozpoczęło przed wojną kilku polskich matematyków z Poznania we współpracy z wywiadami polskim, francuskim i angielskim, ale po upadku Francji inicjatywę przejęli całkowicie Anglicy.

Wyrażenie *artificial intelligence* wymyślił John McCarthy w 1956, jako temat konferencji, która miała nadać, i w rzeczy samej nadała, nowy impet badaniom i projektom prowadzonym w USA. Gdy już mamy ten termin, rodzi się intrygujące filozoficznie pytanie: czy zakresy określeń „sztuczna inteligencja” i „mechaniczna inteligencja” należy uznać za identyczne?

Odpowiedź twierdząca implikuje, że inteligencja wytworzona przez człowieka, a zatem sztuczna (w odróżnieniu od naturalnej) musi być mechaniczna. Czy istotnie musi? Czy sztuczna inteligencja nie może być organiczna? A może dopiero organiczna spełni do końca nasze oczekiwania? Powóćmy do tych pytań pod sam koniec książki, a przygotowują do ich podjęcia dalsze rozważania tego rozdziału.

Po co ludziom inteligentnym wiedza o SI?

Pojęcie człowieka inteligentnego kryje w sobie pewną relatywizację do epoki. W pełni inteligentny jaskiniowiec przeniesiony w naszą epokę miałby nadal znamiona inteligencji, ale nie wszystkie. Wszak inteligencja jest to zdolność rozwiązywania problemów, przed którymi się staje. A ta zależy nie tylko od biologicznych potencji mózgu, którymi mógłby się wykazać i jaskiniowiec, lecz także od sterującego umysłem *systemu pojęć*. Ten zaś musi być dostosowany do miejsca i czasu, w jakim przychodzi człowiekowi rozwiązywać jego problemy, inne dzisiaj niż w paleolicie, inne w wielkiej metropolii niż w pierwotnej dżungli.

Pojęciowy system *społeczeństwa otwartego* które swym zasięgiem pokrywa się mniej więcej z cywilizacją zachodnią, zawiera na poczesnym miejscu pojęcie inteligencji.

Nie chodzi tu o inteligencję w sensie socjologicznym, czyli oświecony ogół, ale o cechę umysłu polegającą na zdolności rozwiązywania problemów. Zakresy terminów „inteligent” i „człowiek inteligentny” nie muszą być identyczne. Powinna tu jednak zachodzić jakaś wysoka korelacja, gdyż istotnym czynnikiem inteligencji jest system pojęciowy; w taki dobrze rozwinięty system mają wyposażać wyższe uczelnie, dające promocję do statusu inteligenta. W tym to systemie znajduje się dziś pojęcie sztucznej inteligencji.

Dlaczego cecha inteligencji znalazła się w centrum uwagi tego typu społeczeństwa, które – za Karlem Popperem – przyjęło się nazywać otwartym? Zasadność pytania jawi się jeszcze wyraźniej, gdy zważyć, że nie znają tego pojęcia wcześniejsze epoki naszej kultury. Gdy znajdujemy, na przykład, w scholastycznej łacinie termin *intelligentia*, to ma on wąski sens specjalistyczny, oznaczając zdolność pojmowania pierwszych zasad filozoficznych (nota bene, mająca w szczególności wysokim stopniu cechować anioły).

Odpowiedzi na to pytanie, dopełniających się wzajem, jest co najmniej cztery. Ich wyliczenie będzie czymś w rodzaju definicji społeczeństwa otwartego. Oto one:

- rola nauki, techniki i oświaty dla zachowania i rozwoju społeczeństwa, które te dobra nauczyło się wytwarzać, a do których kultywowania konieczna jest cecha inteligencji;
- postępująca złożoność życia społecznego wymagająca rozwiązywania coraz większej ilości problemów;
- demokratyzacja społeczna, dzięki której drogą do sukcesu i awansu nie jest odziedziczona pozycja, jak np. w feudalizmie, lecz osobiste walory umysłowe;
- demokratyzacja polityczna, dzięki której metodą walki politycznej nie jest podstęp czy skrytobójstwo lecz publiczna debata wymagająca logicznej argumentacji.

Oczywiście, dwa ostatnie punkty kreślą stan idealny, do którego bywa daleko, jest jednak postęp w tym kierunku.

Skoro pojęcie inteligencji odgrywa tak dużą rolę w myśleniu społeczeństwa otwartego, to żeby odnosić w nim sukcesy trzeba dobrze rozumieć, czym jest owa inteligencja, tak wysoko w nim ceniona. Do niedawna monopol na jej definiowanie mieli psychologowie, ale wraz z rozwojem SI punkt ciężkości przesuwa się w stronę logiki. Widać to w samej psychologii, zwłaszcza w jej nurcie określanym jako *psychologia kognitywna* (za ang. *cognitive psychology*) oparta na pojęciu *przetwarzania informacji*, którego pionierem jest *logika* — nauka o przetwarzaniu informacji pod kątem zachowania jej prawdziwości.

Czy dla zrozumienia natury inteligencji jest potrzebna wiedza o SI? Odpowiedź zdecydowanie twierdząca zawiera się w zdaniu, które widnieje na początku książki P. H. Winstona *Artificial Intelligence*. W oryginale brzmi ono tak: *making computers intelligent helps us understand intelligence*. Oddajmy je przekładem:

**gdy komputer czynimy inteligentnym, pomaga nam to zrozumieć,
na czym polega inteligencja.**

Lekcja z meczu Kasparowa z komputerem

Biegłość w szachach uchodzi za dobry sprawdzian inteligencji. Jeśli więc posiadzie ją maszyna zbudowana sztuką człowieka, można mówić o sukcesie SI.

Dwa pojedynki szachowe Gari Kasparowa z komputerem skonstruowanym przez firmę IBM, pierwszy wygrany przez człowieka, a drugi przez maszynę, śledzone były przez świat z napięciem. Dlaczego? Szukając na to odpowiedzi, dotrzemy do sedna tej myśli, że robiąc komputer inteligentnym czyni się postępy w rozumieniu inteligencji.

Filozoficzny niepokój, który przeżywalismy obserwując zmagania wybitnego umysłu z maszyną wiąże się z jednym z tych powodów, dla których ludzie inteligentni powinni się orientować w SI. Przypomnijmy pewien charakterystyczny komentarz po pierwszej rundzie, tej wygranej przez człowieka.

«17 lutego 1996 roku Gari Kasparow ograł w szachy „Deep Blue”, superkomputer IBM budowany przez pięć lat, wyposażony w monstrualny program. Maszyna ta, kosztująca dwa i pół miliona dolarów, jest w stanie przeprowadzić wyczerpującą analizę wszystkich konfiguracji na osiem ruchów naprzód, a to oznacza około 50 miliardów pozycji. Po porażce „Deep Blue” programiści mieli nietęgę minę, bowiem jako ludzie inteligentni musieli zadać pytanie o inteligencję: Czy inteligencja, błyskotliwość, świadomość, czy te rzeczy w ogóle należą do sfery kalkulacji, obliczeń? Casus Kasparowa unaoczniał raz jeszcze, że nawet najpotężniejszy na świecie komputer nie jest niczym więcej niż kalkulatorem; cóż z tego, że najlepszym na świecie, skoro nawet nie ociera się o inteligencję?» Tak napisał Krzysztof Kowalski w *Rzeczpospolitej* (19-20.XI.1996, s. 20) w rubryce „U progu XXI wieku” w tekście pt. „Świat według Penrose’a”.

Trudno nie ulec pokusie sparafrazowania tego tekstu w taki oto sposób.

porażce Kasparowa z „Deeper Blue” w 1997 roku Krzysztof Kowalski miał nietęgą minę, bowiem jako człowiek inteligentny musiał zadać pytanie o inteligencję. Już wiadomo, że cechujące Kasparowa inteligencja, błyskotliwość, świadomość nie wystarczają do wygrania partii szachów z komputerem.

Już poza tą parafrazą, by autorowi nie imputować refleksji, na którą nie wiadomo, czy by się zdobył, a wiedząc, że wymienione cechy ludzkie nie są do wygrania w szachy z maszyną *wystarczające*, zadajmy teraz zasadnicze pytanie: a czy są one *konieczne*?

I ta odpowiedź będzie negatywna, jeśli błyskotliwości, świadomości etc., nie zamierzamy przypisać maszynie. Na czym zatem polega błąd Krzysztofa Kowalskiego? Na tym, że pochopnie wykluczył grę w szachy „ze sfery – jak powiada – kalkulacji, obliczeń”.

To, że jest to błąd, jeszcze nie przesądza, czy istnieje w ludzkim myśleniu sfera inna niż kalkulacji. Przecież świat myśli nie musi kończyć się na szachach. Ale jeśli ktoś jest przekonany o istnieniu tej innej, to musi się rozejrzeć za lepszymi racjami, niż te, które recytował przytoczony publicysta.

Nie ma po przegranej Kasparowa powodów do filozoficznego niepokoju, czy ludzka inteligencja sprostą maszynowej. Przeciwnie, im lepsza będzie maszyna w rozwiązywaniu problemów dla nas zbyt trudnych, tym lepiej. Budujemy maszyny właśnie po to, żeby robiły mnóstwo rzeczy lepiej niż potrafimy robić to sami. Dotyczy to wielu wymyślonych przez ludzi gier.

Gry takie jak warcaby czy szachy są z natury obiektami matematycznymi, w których wzorcowo się realizują możliwości SI jako potęgi obliczeniowej. Toteż od początków SI należały one, wraz z automatycznym rozumowaniem, do głównych przedmiotów badań oraz projektów technicznych.

I w tej materii pierwszy był Turing. W 1953 opublikował on tekst o zastosowaniu maszyn cyfrowych do gier (*Digital Computers Applied to Games*) w pracy zbiorowej pod znamienym tytułem *Faster than Thought*. Wyrażał tam pogląd, że *szachy* stanowią idealne laboratorium dla SI. Dostarczają one dobrze określonego pola do zgłębiania takich problemów, jak planowanie działań, rola wykorzystywanej w nich wiedzy oraz *poszukiwania heurystyczne* (tzn. idące w przemyślany sposób „na skróty”, gdy pełna realizacja algorytmu jest w praktyce niemożliwa). Jeszcze jedna ich zaleta polega na tym, że dostarczają dobrych miar ilościowych do oszacowania skuteczności różnych technik SI.

Jest to studium, które nakreśliło na ponad dwie dekady kierunki badań nad SI. W szczególności jest jego zasługą określenie funkcji, która oszacowuje w sposób ilościowy siłę lub słabość poszczególnych pozycji w grze. Obecnie programy gry w szachy osiągnęły poziom, który trudno było nawet przewidzieć w tamtym pionierskim okresie sprzed ćwierć wieku, realizującym praktycznie teoretyczne wyniki Turinga. Dowodem ostatni sukces szachowy komputera.

Fakt redukcji myślenia do obliczeń, który jest podstawą SI, wcale nie odbiera tej jakości myśli, którą cytowany wyżej publicysta określił jako *błyskotliwość*, upatrując w niej cechę specyficzną ludzką. Matematyka jest sama w sobie, by tak rzec, głęboko błyskotliwa, to znaczy zdolna zaskoczyć nieoczekiwanym a wielce trafnym wynikiem; po to ją przecież mamy. Gdy matematyk uruchamia w swym mózgu proces rozwiązywania problemu, to uzyskany wynik, nieraz głęboki i błyskotliwy, zawdzięcza on nie tylko własnemu talentowi i wysiłkom, lecz także właściwościom samej matematyki. Dlaczego te same właściwości nie miałyby wyjaśniać sukcesów urządzenia elektronicznego, w które nasza matematyka została wbudowana?

Przekonującego przykładu takiego sukcesu dostarcza opisywany w literaturze przypadek meczu warcabowego z pierwotnej fazy SI, bo z roku 1962. Wystąpił w nim dobrze się w on czas sprawiający program warcabowy A. L. Samuela i bardzo silny gracz warcabowy Robert W. Nealey. Przegrał ją Nealey, a oto jak opisuje on przebieg gry i swoje z niej wrażenia.

«Unikałem posunięć najczęściej podawanych w książkach, bezskutecznie próbując sprowadzić maszynę na nie znane jej tory. O ile mogłem sprawdzić, poczynając od przegrywającego 32 posunięcia, cała reszta partii jest całkowicie oryginalna. Wydaje mi się godne podkreślenia, że maszyna, aby uzyskać zwycięstwo, musiała wykonać kilka błyskotliwych posunięć i że gdyby ich nie zrobiła, osiągnąłbym remis. Dlatego też przedłużałem grę. Jednakże maszyna rozegrała zakończenie partii bezbłędnie. Jeśli chodzi o końcówkę, to nie spotkałem wśród ludzi tak silnego partnera od 1954 roku, kiedy po raz ostatni przegrałem partię.» Są to uwagi podane w *IBM Research News*, a cytowane w pracy zbiorowej *Computers and Thought* pod red. Feigenbauma i Feldmana, 1963.

Mając na uwadze te przykłady, powróćmy do naszej myśli przewodniej, że wyposażając komputer w inteligencję lepiej pojmujemy jej istotę, a na tym tle wyraźniej się zaznacza swoistość inteligencji ludzkiej. Należy do owej istoty zdolność rozwiązywania problemów ujętych w postaci matematycznej, co nie wyklucza rozwiązań, które odczuwamy jako pomysłowe czy błyskotliwe. W tej konkurencji komputer ma szanse konkurowania z człowiekiem.

A co do przewag płynących z posiadania świadomości, to jak długo będzie się utrzymywał prymat człowieka? Czy skończy się on za wiek czy za dziesięć? A może jednak okaże się nie do podważenia? Warto stawiać te pytania, nawet gdy nie stać nas narazie na więcej, niż przegląd stanowisk i wysuwanie hipotez.

Maszyna a świadomość: poglądy A, B, C, D

Faktem jest, że doczekaliśmy się sztucznej inteligencji w sensie powstania programów komputerowych, które dorównują ludziom w rozwiązywaniu problemów. Choćby takich, uchodzących dotąd za niedostępne dla maszyn, jak dowodzenie twierdzeń, gra w szachy, stawianie diagnoz, przekład z jednego języka na inny.

Czy czeka nas i to zaskoczenie, że *powstaną maszyny obdarzone świadomością*? Jest to fundamentalna kwestia filozoficzna. Także technologiczna, bo świadomość jako cecha maszyn rozszerzyłaby kolosalnie ich zdolność rozwiązywania problemów. Byłaby też warunkiem, koniecznym i zarazem dostatecznym, dla partnerstwa maszyn i ludzi we wspólnym myśleniu.

Odpowiedzi na tę kwestię są, oczywiście, dwie – twierdząca i przecząca – ale ten dwuczłonowy podział nie oddaje złożoności problemu. Według Rogera Penrose'a (w *Shadows of the Mind*, 1994) są w tej materii cztery stanowiska, które odróżnia on literami od **A** do **D**. Przyswoimy tę jego charakterystykę, uzupełniając ją o stanowisko (L_m), na tyle ważne filozoficznie, że poświęcimy mu osobny odcinek.

A jest symbolem użytym przez Penrose'a na oznaczenie poglądu znanego szerzej pod nazwą *hard AI* lub *strong AI*. Trudno ją oddać po polsku, bo termin *AI* określa nie tylko samą inteligencję, lecz także prowadzone nad nią badania, czy dotyczące jej teorie. W pojęciu *strong AI* chodzi nie o moc inteligencji, lecz o moc czyli śmiałość, radykalizm, poglądu, że *umysł to nic innego, jak program dla maszyny cyfrowej, a gdy jest to program dostatecznie zaawansowany, to wytwarza świadomość*.

Aksjomatem tej teorii jest pogląd, że *wszelkie myślenie jest obliczaniem*. Tak więc, żeby powstawały akty świadome wystarczy wykonywanie odpowiednich obliczeń. Za inicjatora i filar tego poglądu uchodzi sam Alan Turing.

Stosunek do teorii **A** silnie polaryzuje środowisko badaczy SI. Jedni uważają ten pogląd za jedyny zgodny z postawą naukową, inni traktują go jak nonsens nie zasługujący na uwagę. Jednocześnie zachodzi daleko idące zróżnicowanie wśród jego zwolenników. Trudno, by nie zachodziło, skoro jest więcej niż jedno pojęcie obliczania i więcej niż jedno pojęcie świadomości.

W spektrum odmian **A** spotykamy radykalny pogląd, że wszechświat jako całość jest, w gruncie rzeczy, gigantycznym komputerem. Wśród wykonywanych przezeń obliczeń są takie, które produkują świadomość i świadome siebie samych umysły. Dochodzą do głosu w tym poglądzie trendy i doświadczenia najnowszej nauki, przetworzone na wizję kosmologiczną.

Jednym z nich jest przekonanie, że przedmioty fizyczne redukują się do abstrakcyjnych struktur informatycznych (Penrose nazywa je *patterns of information*). Owe struktury podlegają prawom matematycznych obliczeń, do których się ostatecznie redukują prawa fizyki. Gdy szukać analogii dla tej wizji w dziejach filozofii, to odnajdziemy ją w nurtach pitagorejskich i platońskich, a więc spirytualistycznych, co jest o tyle osobliwe, że mocna SI uchodzi za produkt twardego materializmu.

Inne wszechobecne dziś doświadczenie w uprawianiu nauki stanowią *symulacje komputerowe* oraz produkowanie *rzeczywistości wirtualnej* (te dwa pola zdają się nawzajem przecinać). Doświadcza się w nich stwórczej mocy procesów obliczeniowych, zdolnych wytwarzać nowe realności. Na poziomie zaś technologii komputerowej odpowiada temu częste emulowanie sprzętu przez oprogramowanie. Sa to jakby doświadczalne poszlaki na to, że wszelka rzeczywistość, a więc i umysłowa, może powstawać z procesów obliczeniowych.

A do tego, sama materia jawi się w kontekście współczesnej fizyki nie jako ostoja trwałości i stabilności, lecz jako coś, co da się przemieniać w coś innego, a więc jakby podlegać dematerializacji. Słynne równanie Einsteina mówi o zamianie masy w energię, co toruje drogę myśleniu, że możliwa również jest jej zamiana w coś, co miałoby realność w rodzaju obiektów matematycznych. Sprzyja takiemu myśleniu np. mówienie w teorii kwantów, że cząstki materialne są „falami” informacji.

B jest poglądem podobnym do **A** co do maksymalizmu w ocenie możliwości SI: maszyna może nie tylko dorównać pod każdym względem człowiekowi, lecz nawet pod niejednym go przewyższyć. Różnica w stosunku do **A** wyraża się w tym, że wyprodukowanie świadomości przez procesy obliczeniowe nie jest uważane ani za możliwe ani za konieczne. Nie jest możliwe, ponieważ świadomość jest atrybutem fizjologicznej aktywności mózgu, która nie może zachodzić w urządzeniu elektronicznym. A nie jest konieczne, ponieważ dokładnie te same wyniki można osiągnąć przez symulację świadomości. *Symulacja* jest to matematyczna reprezentacja sytuacji fizycznych i dotyczących ich problemów pozwalająca rozwiązywać te problemy równie skutecznie, jak gdyby zachodziła realnie odpowiednia sytuacja fizyczna. Na przykład, symulacja lotu kosmicznego stawia przed pilotem te same problemy, z którymi miałby do czynienia w rzeczywistym locie.

Idea symulacji może się łączyć z podejściem operacjonistycznym. *Operacjonizm*, kierunek metodologiczny obejmujący teorię definiowania, postuluje definiowanie badanego obiektu nie jako „rzeczy samej w sobie” lecz na określaniu go poprzez sposób działania (operacje). Jeśli uda się wytworzyć robota, który pod każdym względem będzie się zachowywał jak istota, której zachowania tłumaczymy posiadaniem przez nią świadomości, to nie pozostanie nic innego, jak przypisać temu robotowi świadomość.

Stanowisko **B** może przybrać wersję metodologiczną, nacechowaną pewnym agnostycyzmem. Rezygnuje się z definicji świadomości, ponieważ uznaje się ją za zjawisko przekraczające możliwości ścisłego ujęcia naukowego, a zarazem dające się zignorować bez szkody dla badań nad SI.

Czymś w tym rodzaju jest podejście Turinga w słynnym artykule z roku 1950 (zawierającym opis testu nazwanego potem jego imieniem), gdzie pisze on, że wobec niemożności dobrego zdefiniowania, co znaczy „myśleć”, należy określić to pojęcie przez opis zachowań zewnętrznych, które będą nieodróżnialne od zachowań traktowanych jako przejawy stanów wewnętrznych zaliczanych do świadomego myślenia (por. rozdz. 3, odc. 5, fragm. T).

Dalej idąca wersja czyniłaby ze świadomości element będący jakimś ubocznym i nieistotnym produktem ewolucji. Skoro bez świadomości inteligencja może osiągnąć poziom taki sam jak przy jej funkcjonowaniu, a nawet jeszcze wyższy, to jest owa świadomość jakimś przejściowym rozwiązaniem ewolucyjnym, które ustąpi miejsca rozwiązaniom doskonalszym.

Stanowiska **A** i **B** mają to wspólne, że myślenie jest w nich traktowane jako proces zachodzący w układzie fizycznym oraz w pełni algorytmiczny. Przedstawiane dalej poglądy odchylają się od tego wspólnego trzonu stanowisk **A** i **B** w jednym lub w drugim z tych punktów.

C jest poglądem, który ma wspólne z **A** i **B** przekonanie o fizyczności procesów myślenia, ale nie podziela przekonania o ich charakterze algorytmicznym.

Pojęcie *algorytmu* stało się w dobie powszechnego używania komputerów zrozumiałe dla ogółu, przynajmniej w sposób intuicyjny, dzięki pojęciu programu komputerowego. *Program* bowiem jest to algorytm wyrażony w języku przekładalnym na kod maszynowy danego typu komputera. To intuicyjne zrozumienie powinno wystarczyć do śledzenia obecnych wywodów (pełniejsze omówienie pojęcia algorytmu znajduje się w rozdziale 3).

Wymownym i kompetentnym rzecznikiem poglądu **C** jest fizyk angielski Roger Penrose, specjalista od teorii kwantów (znany z osiągnięć w badaniu czarnych dziur). Z teorii kwantów czerpie Penrose przesłanki dla hipotezy o nie-algorytmicznym charakterze procesów mózgowych. Wyraża ją on (w *Shadows of the Mind*, odc. 1.3), jak następuje. *Pewien rodzaj fizycznego działania mózgu prowadzi do powstania świadomości, ale to działanie fizyczne nie da się należycie symulować w sposób algorytmiczny.*

Co do autorstwa tej frapującej hipotezy, trudno je definitywnie ustalić. Pierwsze znane mi prace Penrose'a wyrażające ten pogląd pochodzą z połowy lat 80-tych, ale jego załączkowe sformułowanie pojawiło się 1977 w głośnej książce *The Self and Its Brain*, 1977, której autorami są dwie znakomitości Karl Popper i John Eccles (laureat Nobla 1963 w neurofizjologii). Popper żywił wielkie zainteresowanie teorią kwantów (prowadził na ten temat dyskusje z Einsteinem i wiązał z kwantami swe koncepcje indeterministyczne). Warto tę rzecz mieć na uwadze nie tylko ze względu na kwestię priorytetu, lecz także dla odmienności kontekstów filozoficznych. Popper i Eccles zajmują stanowisko bliższe referowanemu dalej pogładowi **D**, co nasuwa pytanie, na ile i w jakiej wersji podejście kwantowe jest nieodłączne od poglądu **C**.

Penrose sądzi, że działania nie-algorytmiczne mają miejsce w przyrodzie gdzieś poza zasięgiem znanych dotychczas praw fizyki, ale już obecna fizyka dostarcza poszlak na rzecz tej hipotezy. Bierze on pod uwagę i taką wersję **C**, wedle której pewne nowe działy obecnej fizyki, w szczególności teoria *chaosu*, tj. procesów przyrodniczych nieobliczalnych, dotyczyłyby działań nie-algorytmicznych; sam się jednak za tą wersją nie opowiada, stawiając na oczekiwane przyszłe odkrycia nowych praw.

D jest ostatnim punktem w sekwencji uporządkowanej wedle tego, jak wielki wpływ na procesy umysłowe przypisuje się algorytmom. Stanowisko **A** jest najbardziej w tym względzie radykalne, przypisując procedurze algorytmicznej moc wytwarzania świadomości. Na przeciwnym krańcu jest pogląd, że świadomości nie da się wyjaśnić w kategoriach ani algorytmicznych ani fizykalnych.

Jeżeli zamiast „świadomość” powiemy „umysł” czy „dusza”, będziemy mieć odwieczny pogląd, silnie związany z wierzeniami religijnymi; jest nim przekonanie o istnieniu odrębnego od ciała jestestwa duchowego, któremu zawdzięczamy świadomość, myślenie, inteligencję. Niezależnie od związku z religią i teologią, pogląd ten, począwszy od wieku 18-go, cieszy się też wigorem i wzięciem filozoficznym za sprawą René Descartesa, który dał atrakcyjną filozoficznie wizję stosunku między materią i umysłem. Podpisuje się pod nią obecnie tak wpływowy obóz filozoficzny jak fenomenologia, a z innych jeszcze pozycji dostrzegają w niej racjonalne jądro wspomniani wyżej Popper i Eccles.

To wszystko jednak mogłoby nie wystarczyć do tego, żeby badacze SI traktowali tę opcję z równą uwagą jak poprzednie z naszej sekwencji. Swój awans do tego towarzystwa zawdzięcza ona faktowi, który wiąże się z wielkim nazwiskiem austriackiego logika i matematyka Kurta Gödla (1906-1978), czynnego najpierw w Wiedniu, potem w Princeton.

Gödel opublikował na początku lat 30-tych odkrycia, które okazały się przełomowe dla logiki, matematyki, informatyki, filozofii umysłu i filozofii nauki (traktuje o tym szerzej rozdz. 4). Ich znaczenie naukowe jest kolosalne i bezsporne, choć ich interpretacja filozoficzna do dziś kontrowersyjna. Jedną z interpretacji polega na dostrzeganiu w nich argumentu na rzecz stanowiska **D**, a co nadaje jej wagę, to fakt, że podpisał się pod nią sam Gödel.

W logiczno-matematycznej warstwie swych dociekań, wykazał Gödel, że nie może istnieć algorytm dowodzenia twierdzeń w *arytmetyce* — przy założeniu, że jest ona systemem niesprzecznym. Innymi słowy, nie można podać procedury tak ogólnej, by mieć pewność, że przy danym układzie aksjomatów da się podać *dowód* każdego prawdziwego twierdzenia arytmetyki.

„Dowód” jest tu terminem technicznym oznaczającym pewien ciąg przekształceń jednych ciągów symboli w inne, przy czym symbole są traktowane jako przedmioty fizyczne, a więc ich przekształcanie jest procesem fizycznym, wykonalnym dla maszyny wyposażonej w odpowiedni algorytm. Wchodzącymi tu w grę ciągami symboli są zapisy aksjomatów oraz wyprowadzanych z nich twierdzeń.

Jeśli zubożymy *aksjomatykę* o nowe zdania, to dzięki temu więcej da się z niej wyprowadzić, ale pozostaną wtedy inne zdania niedowodliwe a jednak prawdziwe. Wywód ten ma sens przy założeniu, że da się zdefiniować pojęcie *prawdy*, co było przez niektórych odrzucane (nawet z pasją, trudną dziś do pojęcia). Tak się jednak złożyło, że w tymże czasie polski logik Alfred Tarski ogłosił wynik tyleż rewelacyjny, co odkrycia Gödla, mianowicie precyzyjną definicję prawdy dla pewnych języków matematycznych. To pozwala uchylić dawniejszy zarzut (formułowany jeszcze przez biblijnego Piłata), że mówiąc o prawdzie niezbyt wiemy, o czym mówimy.

Pozostał jednak jeden punkt filozoficznie sporny, w którym rozbiegły się drogi Gödla i Turinga, choć ten drugi w parę lat później potwierdził i uogólnił wyniki Gödla w ich warstwie matematycznej (i mającej fundamentalną doniosłość dla informatyki). Można jednak na te wyniki spojrzeć na dwa sposoby.

Gödel, a także Penrose, widzą to w ten podobny sposób (nim ich widzenia nie zaczną się różnić dalej), że matematyk nie będący maszyną ma zdolność poznania prawd nie dających się wywieść z aksjomatów, podczas gdy nie potrafi tego maszyna sterowana algorytmem. Argument na to czerpie każdy z własnych doświadczeń, które przeżywa jako matematyk. Jest to więc argumentacja siłą rzeczy subiektywna, i tu rozeszły się drogi Gödla i Turinga. Turing, owszem, głosił własny i Gödla wynik, że istnieją w matematyce prawdy, do których nie da się dojść algorytmem. Nie sądził jednak, żeby były one bardziej osiągalne dla człowieka niż są dla maszyny. Głęboko był przekonany, że ludzie i maszyny „jadą na jednym wozie”.

Drogi Gödla i Penrose’a rozchodzą się w innym punkcie, mianowicie w pojmowaniu świata fizycznego. Jeśli przyjąć za aksjomat, że świat fizyczny podlega prawom działającym z mocą algorytmu, to możliwość poznawania prawd w sposób inny niż algorytmiczny wymaga przyjęcia jakiegoś czynnika innego niż fizyczny. Taki jest tok myśli Gödla, który sądził, że materialny mózg jest maszyną i podlega prawom algorytmicznym, a wobec tego zdolność poznania prawdy poza wszelkim algorytmem bierze się z jakiejś innej sfery.

Jeśli natomiast dopuści się działania inne niż algorytmiczne w łonie samej materialnej przyrody, to założenie co do istnienia niematerialnego umysłu staje się w tym kontekście zbędne. Taki jest tok myśli Penrose’a.

Czy w powyższych czterech poglądach, uporządkowanych według roli przypisywanej algorytmom, wyczerpują się wszelkie możliwe rozwiązania problemu „maszyna a świadomość”? Z pewnością nie. W każdym razie, istnieje jeszcze jeden wielki pomysł filozoficzny, pochodzący od prekursora idei algorytmu i maszyn cyfrowych, Gottfrieda Wilhelma Leibniza (1646-1716). Poświęćmy mu osobny odcinek, zamykający ten rozdział.

Maszyna a świadomość: pogląd L_m

Wielkość Leibniza jest tak uznana, że z reguły nie ma potrzeby uzasadniania, dlaczego poświęca mu się jeszcze jeden wykład, artykuł czy książkę. Gdy jednak podejmuje się temat tak współczesny, jak SI, to niezależnie od faktu prekursorstwa Leibniza w tej dziedzinie, wypada wyjaśnić, dlaczego się go w to „miesza”. Chylić głowę przed przenikliwością wizjonera to nie znaczy wierzyć, że z tego, co on widział w mglistym zarysie możemy dziś jeszcze czegoś się nauczyć.

Sprawa jednak z Leibnizem jest o tyle nietypowa, że wprowadził on do naszego myślenia o świecie dwie idee tak owocne, że do dziś jest z czego czerpać. Jedną z nich to *Maszyna*, druga to *Nieskończoność*.

Oznaczmy pogląd Leibniza przez L_m – od jego nazwiska; dolny indeks wskazuje na tytuł dzieła, w którym najdobitniej wspomniane myśli wyraził. Jest to *Monadologie* (tytuł oryginału francuskiego), co oddajemy po polsku jako *Monadologia*; jest to zarazem nazwa teorii filozoficznej wyrażonej w tym jego najbardziej dojrzałym dziele (1714). Takie wyróżnienie indeksem jest konieczne, bo – rzecz można – istnieje więcej niż jeden Leibniz (gdy śledzić fazy jego rozwoju) i więcej niż jedna w tej materii interpretacja Leibniza.¹

Maszyna, nazywana też automatem, stanowi centralny punkt zainteresowań Leibniza w *Monadologii*. Jego pojęcie maszyny jest tak szerokie, że obejmuje urządzenia określane tym mianem w języku potocznym, a zarazem obiekty, które dziś byśmy nazwali maszynami informatycznymi.

Te drugie mogą być dziełami człowieka, o czym świadczy choćby maszyna arytmetyczna skonstruowana przez samego Leibniza, ale są wtedy wielce niedoskonałe. Doskonałe maszyny do przetwarzania informacji znajdujemy w przyrodzie, a są nimi *organizmy*. Ale maszyna organiczna to nie jest to, co dziś określamy jako *hardware* (a więc kości, mięśnie etc.), lecz raczej coś, co jest bliższe kategorii *software'u*.

W tym punkcie rysuje się pewne podobieństwo L_m do tego wariantu stanowiska **A**, który można wyczytać u Turinga, a w którym Penrose dostrzega odejście od głoszonych w nim oficjalnie fizykalizmu na rzecz pewnego abstrakcyjnego pojmowania maszyny. Oto jego komentarz.

«Według **A**, *materialna* konstrukcja myślącej maszyny jest uważana za rzecz nieistotną. Tym, co determinuje wszystkie jej atrybuty umysłowe są poprostu wykonywane obliczenia. Obliczenia zaś są same fragmentami abstrakcyjnej matematyki, pozbawionymi wszelkich powiązań z konkretnymi ciałami. Tak więc, według **A**, także atrybuty umysłowe są czymś nie mającym związku z przedmiotami fizycznymi.» (R. Penrose, *Shadows of the Mind*, odc. 1.3, przekł. W.M.).

U Leibniza nie jest aż tak prosto, bo związki umysłu z ciałem nie podlegają takiemu rozdzielaniu jak w powyższym obrazie, ale łączy go z poglądem **A** to, że wyższy poziom rzeczywistości przysługuje umysłom, te zaś są częścią lub, raczej, wytworem abstrakcyjnej matematyki. Ciała są jakby sposobem przejawiania się owych abstrakcyjnych jestestw wobec ludzkich zmysłów.

Czy w poglądzie L_m jest miejsce na sztuczną inteligencję? Nakreślone wyżej pojmowanie organizmów jako abstrakcyjnych maszyn jeszcze nie dostarcza na to odpowiedzi. Odpowiedź przychodzi dopiero po dołączeniu nowej przesłanki, jaką jest Leibnizjański infinityzm — wiara w nieskończoną złożoność organizmów i związek inteligencji z tą nieskończoną złożonością.

Nieskończoność stanowi drugi klucz do Leibnizjańskiej wizji świata. Wprawdzie Leibniz nie zdawał sobie sprawy z tego, że istnieje nieskończenie wiele zbiorów nieskończonych; jego pogląd pozostaje więc niedookreślony co do tego, jaką nieskończoność ma się na myśli. Wystarczy jednak dla naszych rozważań, gdy weźmiemy pod uwagę najmniejszą z nich: taką, jaką cechuje zbiór liczb naturalnych.

Aby odnieść to pojęcie do przyrody tak, jak czynił to Leibniz, trzeba mieć na uwadze, że był on zdecydowanym przeciwnikiem atomizmu. Uważał za niemożliwe, by podział ciał na mniejsze części miał jakikolwiek kres, który to kres wielu jego współczesnych upatrywało w atomach (rozumianych w sensie Demokryta). Głosił więc podzielność ciał w nieskończoność, co znaczy, że dla każdej struktury materialnej istnieją struktury będące jej składnikami.

¹ Wobec wielości interpretacji, litera „m” może być także czytana jako wskaźnik, że jest to propozycja interpretacyjna pochodząca od obecnego autora, podpisującego się swym inicjałem. Taki „podpis” autora interpretacji ma świadczyć, że nie jest ona jedyną możliwą.

Gdy dzielimy w ten sposób ciało ożywione, to jego funkcjonowanie – powiada Leibniz – na każdym poziomie złożoności zależy od funkcjonowania struktur z kolejnego poziomu podziału, i tak w nieskończoność. Z tego, co wiemy dziś na temat organizmów, pogląd ten ma do pewnego punktu klarowną interpretację empiryczną. Wiemy, podobnie jak wiedziano w czasach Leibniza, że funkcjonowanie całego układu żywego zależy od jego podsystemów, jakimi są np. poszczególne organy. Wiemy też, czego przed dwoma wiekami nikt wiedzieć nie mógł, że takie zależności sięgają aż poziomu molekularnego (idea, którą dopiero w roku 1941 rzucił wybitny fizyk atomowy Erwin Schrödinger).

Na tym poziomie zatrzymuje się dziś front oficjalnej biologii, ale oto wybiegają przed linię frontu co zuchwalsi i głoszą – jak Penrose, czy Popper z Ecclesem – że w funkcjonowaniu mózgu zachodzą zależności od głębiej położonych struktur sięgające poziomu kwantowego. Trzeba to uznać za kolejny punkt dla Leibniza, ale do wykazania jego racji jest jeszcze daleko, można rzec, nieskończenie daleko.

Dalej wszystko zależy od tego, czy ma jakiś sens fizyczny mówienie o *nieskończonej złożoności materii*. Jest to warunek konieczny (choć nie wystarczający) przyznania racji Leibnizowi, gdy głosi, że aby zrozumieć bez reszty funkcjonowanie organizmu, w szczególności jego zachowania inteligentne, trzeba by ogarnąć poznawczo cały nieskończony ciąg stopni złożoności.

Nie mogąc w tej sprawie podać definitywnego argumentu można jednak zestawić aktualne opinie fizyków przychylnie wizji Leibniza; nie wystarczy to, by potwierdzić jego racje, ale starczy jako motyw do dyskusowania ich na serio. Zaczniemy od wypowiedzi wybitnego polskiego matematyka Stanisława Ulama, który (jako jedyny z Polaków) uczestniczył w pracach słynnego laboratorium w Los Alamos prowadzących do skonstruowania bomby atomowej. Choć nie jest on typowym przedstawicielem fizyki jak przywoływani potem autorzy, to warto zacząć od jego sformułowań, bo brzmią jakby przepisane, co do głównej myśli, z Leibniza (choć ich autor Leibniza w tej materii nie studiował).

«Według mnie pierwszym pytaniem fizyki – choć, oczywiście, nie można tego uznać za precyzyjnie sformułowany problem – jest to, czy istnieje prawdziwa nieskończoność struktur o coraz mniejszych i mniejszych rozmiarach. Jeśli tak, to matematycy mogliby zastanowić się nad tym, czy czas i przestrzeń nie zmieniają się, może nawet pod względem topologicznym, gdy przechodzimy do coraz mniejszych obszarów. W fizyce istnieją podstawy atomistyczne albo oparte na teorii pola. Jeśli rzeczywistość koniec końców ma charakter polowy, to punkty są prawdziwymi punktami matematycznymi i są nierozróżnialne. Istnieje też możliwość, że mamy do czynienia z dziwną strukturą o nieskończenie wielu poziomach, a każdy z nich ma inną naturę. Jest to nie tylko zagadka filozoficzna, ale i fascynująca coraz bardziej fizyczna wizja. Ostatnie eksperymenty wykazują narastające skomplikowanie struktur. Pojedynczy nukleon może składać się z partonów, jak to nazywa Feynman. Mogą to być hipotetyczne kwarki lub inne obiekty. [...] Być może, osiągnęliśmy punkt, gdzie lepiej byłoby rozważyć następstwo struktur *ad infinitum*.» Stanisław M. Ulam, *Przygody matematyka*, przełożyła A. Górnicka, 1996 (s. 324n); pierwsza publikacja ang. oryginału *Adventures of a Mathematician*, 1976.

Z większą rezerwą, ale także i z respektem dla tej ewentualności, która się zawiera w wizji Leibniza, wypowiada się sławny Stephen W. Hawking (promotorem jego doktoratu był nasz znajomy z tych stron Roger Penrose).

«Nauczyliśmy się ostatnio używania pól elektromagnetycznych do nadawania cząstkom coraz większych energii. Stąd wiemy, że cząstki, które uważano za „elementarne” przed dwudziestu laty są, w gruncie rzeczy, złożone z mniejszych cząstek. Czy może się okazać, w miarę przechodzenia do jeszcze wyższych energii, że i te właśnie odkryte są złożone z jeszcze mniejszych cząstek? Jest to z pewnością możliwe, ale mamy pewne powody teoretyczne, by wierzyć, że zbliżamy się do poznania ostatecznych cegiełek przyrody.» Stephen W. Hawking, *A Brief History of Time from the Big Bang to Black Holes*, 1sze wyd. 1988 (s.66, przekład ad hoc W.M.).

Jeszcze jedna opinia warta jest przytoczenia, choć nie wyszła spod pióra fizyka lecz popularyzatora fizyki; popularyzator to jednak wysoce kompetentny.

«Każda zmiana skali przynosi nowe zjawiska i nowe sposoby zachowania. Dla nowoczesnej fizyki cząstek elementarnych proces ten nie ma końca. Każdy nowy akcelerator, z jego wzrostem energii i prędkości, rozszerza pole widzenia nauki ku coraz mniejszym cząstkom i coraz krótszym skalom czasowym, a każde rozszerzenie przynosi nową informację.» James Gleick, *Chaos*, 1sze wyd. 1988 (s. 115, przekład ad hoc W.M.).

Tekst Ulama ukazuje pewną wizję; dwa następne mówią o warunkach eksperymentalnych niezbędnych dla jej weryfikacji, z czego wynika wniosek o granicy naszego poznania, gdy idzie o docieranie do coraz głębszych struktur. Jest to granica ekperymentalna, bo nie można zwiększać wytwarzanych energii w nieskończoność. Dodajmy do tego zdanie Wernera Heisenberga, jednego z głównych twórców teorii kwantów, mówiące o ograniczeniu naszych możliwości teoretycznych.

«Przechodząc do wielkości coraz mniejszych, nie znajdujemy wielkości fundamentalnych i niepodzielnych, natomiast dochodzimy do punktu, w którym dalsze dzielenie już nie ma sensu.» Cytowane za książką Paula Daviesa *Bóg i nowa fizyka*, 1996, przełożył P. Amsterdamski (rozdz. „Fundamentalna struktura materii”). Wyd. oryginalne *God and the New Physics*, 1983.

Chociaż tok myśli ilustrowany powyższymi cytatami nie jest konkluzywny, to znaczy, nie doprowadza do rozstrzygnięcia jak to jest naprawdę ze stosunkiem materii do inteligencji, dostarcza on rusztowania pojęciowego na tyle, by rozpocząć budowlę, wznieść jakby pierwszą jej kondygnację. Jest nią wyodrębnienie problemów, na które trzeba odpowiedzieć, żeby przystąpić do wznoszenia dalszych kondygnacji. Rusztowaniem zaś jest idea stopni złożoności materii. Pozwala ona postawić pytanie o stosunek między kodem stosowanym do sterowania daną strukturą a jej usytuowaniem na skali poziomów złożoności; na przykład: czy na każdym poziomie stosowny jest kod binarny? Wyjaśnijmy to dokładniej.

Kod sterujący zachowaniem układu jest czymś, z czym myśl ludzka oswaja się stopniowo, poczynając od budowanych już przed paru wiekami automatów. Były to instrumenty muzyczne i różne zabawki, a także maszyny tkackie; środkiem sterowania była, na przykład, dziurkowana taśma, w której konfiguracja dziurek określała zamierzone przez konstruktora zachowania układu (np. wytworzenie określonego wzoru na tkaninie). System tworzenia owych konfiguracji podpada pod nasze dzisiejsze pojęcie *kodu*.²

Na tamtym etapie istnienie kodów było faktem marginesowym, bez większego wpływu na poznawanie i przekształcanie świata. Nikt też poza Leibnizem nie sądził, że kody mogą być decydującym czynnikiem w życiu organizmów. U Leibniza wynikało to z jego koncepcji organizmów jako myślących automatów; nikt jednak, nawet wśród wiernych komentatorów, nie był w stanie docenić tej myśli. Jej czas nastał dopiero wraz z konstruowaniem maszyn cyfrowych i równoległym procesem odkrywania kodów w przyrodzie: genetycznego i neuronowego.

Nie stało się tak bynajmniej w wyniku czytania Leibniza. Ale gdy rozwój nauki doszedł już do tego punktu, to konsultacja u Leibniza powinna pomóc w uświadomieniu, że pojęcia *kodu* i zapisanego w nim programu są kluczem do rozumienia świata. W szczególności, do rozumienia relacji w trójkącie *inteligencja, świadomość, materia*.

W tej perspektywie przyjrzyjmy się raz jeszcze stanowisku oznaczonemu wyżej przez **A**, żeby zadać pytania naprowadzające w wyniku na L_m . Zwolennicy **A**-izmu powiadają, że świadomość, a wraz z nią wejście do elity jestestw inteligentnych, można zapewnić również

² W innym sensie mówimy o kodzie, gdy chodzi o szyfrowanie informacji w celu ich utajnienia. Tutaj mowa jest cały czas o kodach sterujących zachowaniem układów.

urządzeniom elektronicznym. Trzeba je tylko wyposażyć w odpowiednio wyrafinowany program (pesymiści w tym obozie oczekują, że stanie się to około roku 2030, optymiści dają bliższe terminy). Znaczy to, co następuje.

- 1. Wystarczający do wytworzenia świadomości jest stosowany w maszynach cyfrowych kod binarny.
- 2. Struktura materialna wystarczająca do zapisywania w tym kodzie programów znajduje się na poziomie złożoności reprezentowanym przez elektrony.
- 3. Wystarczającym do tego rodzajem materii jest krzem — ten rodzaj kryształu, z którego buduje się obecnie procesory i pamięci maszyn cyfrowych.

Rząd wielkości osiągalny dla technologii wymienionej w 3 można zilustrować faktem, że w układach o najwyższym stopniu scalenia na płatku kryształu o powierzchni 66 mm^2 mieszczą się miliony elementów tego typu, że każdy odpowiada jednej lampie elektronowej z pierwszej generacji maszyn cyfrowych (lata 1946-1959). Powiedzmy, że w którejś kolejnej generacji osiągnie się miliardy czy biliony elementów na jednej płytce. Im większa liczba elementów do zapisu liczb, tym większe liczby można zapisywać za ich pomocą w kodzie binarnym.

Kto jednak wie, jak wielkie muszą to być liczby, żeby za ich pomocą dał się wyrazić w kodzie binarnym program generujący świadomość? Może, na przykład ten jej rodzaj, który nazywamy świadomością moralną czyli sumieniem jest tak subtelny, że liczba zastosowana do jego opisu będzie wymagać nieosiągalnej technologicznie ilości miejsc po przecinku? Może będzie to ilość tak wielka, że potrzebna do tego ilość komórek pamięci nie zmieści się w układach o najbardziej zawrotnej skali integracji?

Ten rodzaj pytań jest inspirowany poglądem Leibniza, że program definiujący żywy organizm wyraża się liczbą rzeczywistą mającą nieskończone i nie dające się obliczyć w skończonym czasie rozwinięcie. Z tego poglądu natychmiast wynika odpowiedź przecząca na pytanie o możliwość wytworzenia, czy choćby adekwatnej symulacji świadomości w jakimkolwiek układzie skończonym. Staje się to, wedle Leibniza, możliwe dopiero w układzie nieskończonym tego typu, że program jest rozpisywany na coraz głębsze poziomy złożoności; a ponieważ jest ich nieskończenie wiele dzięki nieskończonej złożoności strukturalnej (której dotyczył cytat z Ulama), stworzyć nieskończony program może tylko programista o nieskończonej potencji intelektualnej. Nie jest to zatem wykonalne dla człowieka.

To skrajne stanowisko Leibniza spróbujmy uczynić bardziej umiarkowanym, unikając też konkluzji w sprawie nieskończonego umysłu programisty, przez wstrzymanie się od sądu co do tego, czy liczba poziomów złożoności musi być nieskończona. Wtedy pozostaniemy bez konkluzji, ale za to z otwartym bardzo pouczającym pytaniem: jak wielka to ma być liczba, żeby wystarczyła do zapisu programu odpowiadającego swą złożonością skomplikowaniu fenomenu świadomości?

Takie postawienie sprawy pomimo braku konkluzji posuwa naprzód dyskusję nad SI. A posuwa ją dzięki temu, że sytuacja dyskusyjna klaruje się przez wyraźne wskazanie, na której stronie ciąży obowiązek dowodzenia czyli *onus probandi* (jak to się formułuje w logice prawniczej). Kto się zobowiązuje wyposażyć maszynę cyfrową w świadomość, jak to czyni **A**-ista, niech oszacuje liczbowo, jakiego stopnia złożoności będzie wymagał odpowiedni program i jak to się ma do osiągalnego technologicznie stopnia złożoności układów scalonych, a także do czasu potrzebnego na samo programowanie, a potem na realizowanie programu.

A jeśli nasz **A**-ista dojdzie do wniosku, że nie jest to wykonalne dla obecnej technologii, niech zaprojektuje inną. Może to będzie inny rodzaj kodu aniżeli binarny, na przykład jakiś kod analogowy, zdolny adekwatnie odwzorowywać wszelkie liczby rzeczywiste? A może potrafi on zejść na inny, głębszy, poziom złożoności, dostarczający nowych możliwości kodowania?

Pytania takie nieprędko się zapewne doczekają odpowiedzi na poziomie intersubiektywnej dyskusji naukowej. Każdy z nas jednak, zachowując należyty krytycyzm, ma prawo do subiek-

tywnej odpowiedzi, bodaj hipotetycznej, na swój osobisty użytek. Podstawą do takiej odpowiedzi będzie własne doświadczenie wewnętrzne wiodące do oszacowania, jak dalece złożonym fenomenem jest nasza świadomość.

Czynią tak przecież ci rzecznicy poglądu **A**, którzy zapowiadają wyprodukowanie świadomości komputerów, powiedzmy, do roku 2030. Nie dysponują oni obliczeniami, o których wyżej była mowa, ale mają jakieś osobiste poczucie, że świadomość to byt stosunkowo nieskomplikowany, tak więc, szacując „na oko”, sprostą mu technologia spodziewana w najbliższych dekadach. Tak wolno szacować każdemu, byłoby jednak niezupełnie w porządku ogłaszać takie oczekiwania jako obiektywne i jedynie naukowe podejście.

Podobne prawa mają ci, którym własne doświadczenie wewnętrzne podpowiada, że jest w świadomym sobie umyśle głębia złożoności, której nie sprostą kod binarny z technologią elektroniczną. Potrzebną do tego technologię Leibniz nazywał boską — *mechanica divina*. Jeśli ten przymiotnik wyraża poczucie tajemnicy, to znaczy nieskończonych obszarów poznawczych otwierających się przed umysłem, nie muszą się go wyrzekać nawet przysięgli agnostycy. Im też przystoi, by nie twierdzić w jakiegokolwiek sprawie, także w sprawie SI, „będziemy wiedzieć wszystko”; mają natomiast prawo mówić z przekonaniem: „będziemy wiedzieć coraz więcej”.

*

W tej perspektywie, sięgnijmy po odpowiedź na pytanie postawione w tytule: Czy już żyjemy w erze sztucznej inteligencji?

Odpowiedź brzmi NIE, jeśli *sztuczna inteligencja* pojmować jako *inteligencja maszynową wytworzoną przez człowieka oraz równą ludzkiej w sensie posiadania świadomości*. Sztuczność znaczy wtedy tyle, że chodzi o dzieło ludzkie nie zaś o twór przyrody, może to być jednak dzieło dorównujące przyrodzie. Takie zrównanie maszyny z umysłem dotąd nie nastąpiło, a czy może kiedyś nastąpić — oto jest pytanie. Pytanie filozoficzne, na które różni różnie odpowiadają. L_m odpowiada przecząco.

Odpowiedź brzmi TAK, jeśli *sztuczna inteligencja* pojmować jako *zdolność maszyny do wytwarzania produktów naśladujących wiernie pewne wytwory inteligencji ludzkiej* — partie szachowe, dowody twierdzeń, przekłady z języka na język. Jest to „TAK” stopniowalne, bo jesteśmy dopiero u początku drogi. Będziemy się nią posuwać dalej i dalej, w miarę jak będzie wzrastać, do rozmiarów nie dających się dziś wyobrazić, interakcja maszyn i umysłów w jednej globalnej sieci (traktuje o tym rozdz. 8).

Nasz czas jest porą stawiania pierwszych kroków. Należy do nich zrozumienie, na czym polega stosunek między liczeniem i myśleniem. Jest o tym mowa w następnym rozdziale.