

P a w e ł S t a c e w i c z

Czy informatykom uda się skonstruować maszyny autonomiczne?

1. Wstęp

We współczesnym świecie aż roi się od maszyn. Dla wielu osób stały się one nieodzownym elementem życia codziennego, towarzystwem tak częstym i tak pożądanym jak roślina, zwierzęta czy książki. Taki stan rzeczy nie mógłby zaistnieć bez długotrwałego rozwoju cywilizacji technicznej, która wydała z siebie zrazu maszyny do przetwarzania materii i energii, a współcześnie maszyny do przetwarzania informacji. W niniejszej pracy skoncentrujemy się na urządzeniach ostatniego rodzaju.

Ponieważ tekst ma charakter filozoficzny, skupimy się w nim nie na szczegółach technicznych budowy tychże urządzeń, lecz na ogólnej dyskusji o ich perspektywach rozwojowych. Jej ramy wyznaczają dwa stanowiska skrajne. Pogląd pierwszy, który można nazwać *minimalistycznym*, polega na uznawaniu systemów informatycznych za zwykłe narzędzia, a więc skonstruowane przez ludzi urządzenia, które pomagają im realizować rozmaite zadania. Pogląd drugi zaś wyraża *maksymalistyczne* przekonanie o tym, że dzisiejsze układy do przetwarzania informacji stanowią punkt wyjścia do skonstruowania maszyn autonomicznych, a więc niezależnych podmiotów własnych działań, które mogłyby stać się partnerami lub konkurentami ludzi.

Czy w dyskusji nad kształtem i zasadnością powyższych stanowisk powinni uczestniczyć filozofowie, a nie tylko przedstawiciele informatyki? Naszym zdaniem tak, a przemawiają za tym następujące argumenty. Po pierwsze, filozofowie są bardzo dobrze predysponowani do oglądu dyscyplin szczegółowych, w tym ich wytworów, z pewnej perspektywy. Dzięki temu mogą wskazywać w nich pewne cechy ogólne, sprzyjające takim a nie innym kierunkom rozwojowym. Po drugie, filozofia, a zwłaszcza epistemologia, zawsze zajmowała się formułowaniem kryteriów oceny wytworów różnych nauk – kryteriów zewnętrznych wobec metod tychże nauk. Przykładowy i tradycyjny przedmiot tego rodzaju badań

stanowią kryteria prawdziwości twierdzeń matematycznych i przyrodniczych. W odniesieniu do maszyn wynika natomiast zagadnienie nierozróżnialności, to znaczy zagadnienie wyrażające się pytaniem: „na podstawie jakich symptomów, testów czy badań moglibyśmy wnosić, że udało nam się skonstruować urządzenia autonomiczne, a nie – będące wyłącznie naszymi narzędziami?”. Po trzecie wreszcie, pytanie o autonomię konstruowanych systemów informatycznych nawiązuje wprost do dyskusji filozofów o tym, jakie cechy ludzkiego umysłu przesądzają o jego autonomii i czy kiedykolwiek cechy te będzie można modelować. Uczestniczą w niej między innymi epistemologowie, filozofowie umysłu i filozofowie języka.

Na zakończenie wstępu winniśmy odnieść się bliżej do znaczenia zwrotu „maszyna autonomiczna”. Otóż na ogół będziemy go rozumieć wąsko i dość specyficznie jako równoważnik wyrażenia „maszyna niebędąca tylko narzędziem człowieka”, przy czym istotę narzędziowej funkcji maszyn postaramy się wyjaśnić w ustępie 3. Sam wyraz „autonomia” przywodzi jednak na myśl szereg skojarzeń ogólnych, takich jak „niezależność”, „podmiotowość” czy „samoświadomość”. Z tego względu jest użyteczny, bo daje jakieś wyobrażenie o tym, co może się kryć za poglądem dopuszczającym skonstruowanie maszyn niebędących wyłącznie naszymi narzędziami. Kwestiom tym przyjrzymy się bliżej w ustępie 4.

2. Maszyny do przetwarzania informacji

Filozof, który chciałby wyjaśnić, na czym polega istota systemów informatycznych, staje przed nie lada wyzwaniem. Materiał empiryczny, do którego musi się odnieść, jest bowiem niezwykle bogaty i różnorodny. Z jednej strony składają się nań zasady działania systemów już skonstruowanych, takich jak komputery cyfrowe, maszyny analogowe, komputery równoległe i różnego rodzaju sieci komputerowe, z drugiej strony zaś ogromna liczba pomysłów czekających na realizację, chociażby komputerów kwantowych, ewoluujących czy wykorzystujących żywą materię. Rzecz całą utrudnia istny zalew metod przetwarzania danych (np. konkretnych algorytmów), które często tworzy się *ad hoc*, bez odpowiedniej teorii i na użytek konkretnych zastosowań. Najbardziej chyba wymownym przykładem owego prymatu praktyki nad teorią pozostaje brak zadowalającej definicji samej informacji, a więc tego czegoś, co przetwarzają już konstruowane i już użytkowane maszyny.

Zarysowane wyżej trudności traktujemy jako pewne usprawiedliwienie dla treści niniejszego ustępu, w którym zwrócimy uwagę na wybrane tylko cechy systemów informatycznych. Cechy te dobierzemy tak, aby uwypuklić różnice między maszynami tradycyjnymi a maszynami do przetwarzania informacji. Najważniejsza z nich da się wyrazić następująco: *o ile maszyny pierwszego rodzaju przetwarzają materię lub energię, maszyny drugiego rodzaju operują na czystych da-*

nych¹. Na przykład komputer wyposażony w program do sortowania listy słów przekształca dane początkowe, czyli nieuporządkowaną listę odpowiednio zakodowanych słów, w dane końcowe, a więc listę słów uporządkowaną według podanej zasady. Jego działanie wymaga wprawdzie niewielkiej ilości energii oraz pośredniego przetworzenia materii (tj. strumienia elektronów przepływających przez układy scalone), jednak są to tylko czynności pomocnicze względem przekształcania danych. Dla kontrastu obrabiarka – typowa przedstawicielka maszyn tradycyjnych – służy po prostu do obróbki materii. (Dodajmy na marginesie, że wskutek posiadania silnika obrabiarka przetwarza również energię, na przykład elektryczną w mechaniczną, a gdy steruje nią odpowiednio zaprogramowany komputer, przetwarza również dane)².

Przywoływany dwukrotnie przykład komputera cyfrowego – a więc maszyny operującej wyłącznie na zapisach symbolicznych (dyskretnych)³ – pozwoli nam wyjaśnić dokładniej, czym w rzeczywistości mogą być dane. Wiadomo, że specjaliści programujący i obsługujący takie komputery używają w różnych sytuacjach różnych kodów, co przejawia się na przykład w tym, że wybierają różne struktury danych i różne języki programowania. Korzystając z nich, zyskują komfort komunikowania się z maszyną w języku przypominającym nieco język naturalny. Z drugiej strony wiadomo, że dane rzeczywiście przetwarzane – a nie wprowadzane przez człowieka lub dostarczane mu przez maszynę – koduje się w postaci sekwencji zer i jedynek. Decydują o tym względy techniczne, głównie zaś potrzeba zminimalizowania błędów w przekazie danych (przypomnijmy, że jedyne odpowiada obecność sygnału elektrycznego, a zera – jego brak). Podsumowując więc, można powiedzieć, że w przypadku komputera cyfrowego dane przyjmują postać dwojaką: dla użytkowników są to językopodobne i znaczące ciągi symboli, dla maszyny – wyłącznie ciągi zer i jedynek.

Niezależnie od takich czy innych form zapisu dane stanowią tworzywo zupełnie odmienne od materii czy energii. Z pewnego punktu widzenia można je nazwać abstrakcyjnym, bo jest oderwane od konkretnego kodu i jego interpretacji, z innego – reprezentującym, bo nadaje się znakomicie do reprezentowania zew-

¹ Chcąc mówić o tworzywie przetwarzanym przez systemy informatyczne, musimy pozostać na poziomie *danych*, a więc zapisów informacji w pewnym kodzie. Zabieg utożsamienia informacji z danymi jest nieodzowny wobec sygnalizowanego wcześniej braku zadowalającej definicji semantycznej samej *informacji*.

² Pogłębioną analizę różnic między maszynami tradycyjnymi a systemami informatycznymi zawierają książki Bolter, J.D., *Człowiek Turinga. Kultura Zachodu w wieku komputera*, PWN, Warszawa 1998 i Marciszewski, W., *Sztuczna inteligencja*, Znak, Warszawa 1998. W obydwu cytowanych pracach jest ponadto prezentowane stanowisko, zgodnie z którym upowszechnienie się odpowiednich maszyn miało decydujący wpływ na krystalizowanie się trzech ogólnych pojęć teoretycznych: *materii*, *energii* i *informacji*.

³ Odmienny sposób kodowania obowiązuje w przypadku maszyn *analogowych*, które nie operują na zapisach symbolicznych, lecz odwzorowują wielkości (kody) ciągle w inne wielkości ciągle.

nętrnych względem siebie obiektów, stanów i procesów. Obydwie własności danych decydują o znanej ogólnie cesze przetwarzających je maszyn, a mianowicie ich *skrajnej wielofunkcyjności*⁴. Wszak to samo urządzenie fizyczne może być wykorzystywane i do rozwiązywania równań, i do całkowania, i do nauki języków obcych, i do odtwarzania utworów muzycznych. O jego bieżących funkcjach decyduje uruchamianie na nim *oprogramowanie*. Wymienne oprogramowanie należy traktować jako nieodłączny składnik wszelkich maszyn do przetwarzania informacji.

Pośród rozlicznych, uzyskiwanych za pomocą odpowiednich programów, funkcji systemów informatycznych szczególną uwagę zwraca możliwość sterowania maszynami tradycyjnymi. Zostaje ona zrealizowana, gdy dane przetwarzane przez system są wykorzystywane jako impulsy do konkretnych działań maszyny – na przykład, jeśli powołamy się na wcześniejszy przykład obrabiarki, do zmiany nacisku elementu tnącego. Ze względu na owe funkcje kontrolno-sterujące systemów informatycznych można przypisać im status *nadrzędności* wobec maszyn tradycyjnych. Wolno to uczynić tym bardziej, że za ich pomocą daje się symulować pracę dowolnych maszyn tradycyjnych bez ich faktycznego uruchamiania. *De facto* możliwość tę wykorzystuje się nader często, na przykład podczas szkolenia pracowników, którzy będą obsługiwać w przyszłości jakieś maszyny tradycyjne, lub wtedy, gdy trzeba przetestować pewne rozwiązania prototypowe i jeszcze niewdrożone⁵.

Opisywane wyżej cechy miały charakter różnicujący. Ta, o której wspomniemy na koniec, przysługuje również wielu maszynom tradycyjnym. Chodzi o fakt, że systemy informatyczne działają *automatycznie*, to znaczy raz uruchomione – przez człowieka lub inną maszynę – wykonują swoje zdania samodzielnie. Użytkownik może co najwyżej przerwać ich działanie lub, jeśli przewiduje to program maszyny, inaczej je ukierunkować. Zauważmy tu wstępnie, przed przejściem do ustępu 4, że wymóg pracy bez ingerencji człowieka stanowi warunek konieczny przekształcenia maszyn z narzędzi (przedmiotów) w wytwory autonomiczne (podmioty).

Na koniec niniejszego ustępu zadajmy pytanie, które odnosi nas wprost do związków między systemami informatycznymi a ludźmi. Gdzie znajduje się siedlisko naszej inteligencji, woli, świadomości i wszystkich tych dyspozycji, które czynią nas podmiotami własnych działań, czyli – nawiązując do tytułu niniejszej

⁴ Słowa „skrajnie” używamy po to, by podkreślić odmiennność systemów informatycznych od maszyn tradycyjnych, których budowa i realizowane funkcje są dość ściśle zdeterminowane własnościami przetwarzanego tworzywa (choć często maszyny takie też są wielofunkcyjne).

⁵ W odniesieniu do przekształcanego tworzywa nadrzędność maszyn do przetwarzania informacji wobec maszyn tradycyjnych przejawia się w tym, że mimo bezpośredniego przetwarzania informacji mogą służyć one pośrednio do przetwarzania materii lub energii. Pośrednio, bo wskutek odpowiedniego zinterpretowania i wykorzystania przetworzonych danych.

pracy – decydują o naszej autonomii? Odpowiedź właściwa filozofii brzmiałaby krótko: *w umyśle*. Odpowiedź, której udzieliłby biolog, miałaby postać bardziej uchwytnej: *w mózgu*. A jakie funkcje odróżniają mózg od reszty organów? Czy nie to, że kontroluje on ich pracę? Czy nie to, że operuje wyłącznie na odpowiednio zakodowanych danych zmysłowych? Czy jego typowo ludzkie funkcje nie wykształciły się w toku ewolucji najpóźniej? Pamiętamy jednak, że trzy podobne własności – przetwarzanie danych, nadrzędność wobec innych maszyn i „najkrótszy cywilizacyjny staż” – uznaliśmy za wyróżnik systemów informatycznych. Czy owa zaskakująca analogia między mózgami a systemami informatycznymi nie powinna skłonić nas do wzięcia na serio pytania postawionego w tytule niniejszej pracy?

3. Systemy informatyczne jako narzędzia

Skoro wiemy już, na czym polega specyfika współczesnych maszyn do przetwarzania informacji i jakie pobudki skłaniają niektórych do uznawania ich za łańczki autonomicznych maszyn przyszłości, przyjrzyjmy się bliżej postawie innej, którą nazwaliśmy we wstępie minimalistyczną. Postawa ta jest chyba najbardziej zdroworozsądkowa i ugruntowana w teraźniejszych dokonaniach informatyki, a polega na przyznawaniu systemom informatycznym statusu zwykłych *narzędzi*.

Narzędzia te mają szereg cech swoistych i są od innych bardziej wyszukane, niemniej pozostają tylko i wyłącznie użytecznymi przedmiotami w ręku ich twórców. A zatem: „Pewien program komputerowy wykonuje automatycznie jakieś operacje, na przykład dowodzi twierdzenia matematyczne, lecz to do nas należy interpretacja jego wyników”, „Pewien system pracuje bez naszego nadzoru, ale to my go skonstruowaliśmy i zawsze możemy zmienić jego sposób działania”. Innymi słowy, systemy informatyczne nie działają same z siebie, lecz to my działamy za ich pośrednictwem. Konsekwentni zwolennicy omawianego poglądu podkreślają, że stosowanie do maszyn określeń antropomorfizujących zawsze będzie miało charakter czysto metaforyczny. Aby rzecz unaocznić i odnieść do systemów informatycznych, odwołajmy się do porównania. Tak jak powiemy, że „to nie kalkulator szybko liczy, a my szybko liczymy, posługując się nim”, tak musimy orzec, że „to nie komputer przetwarza informacje, ale my za jego pomocą”.

Pogląd scharakteryzowany wyżej bardzo ogólnie może występować w wielu odmianach, z których nie wszystkie wydają się tak zdroworozsądkowe jak ogólne hasło „maszyna to zwykle narzędzie”. Przede wszystkim wyznając go, można myśleć o *narzędziach praktycznych*, a więc takich, które pozwalają nam osiągać pewne konkretne cele, np. przechowywać dane, sterować ruchami robotów lub rozwiązywać skomplikowane równania matematyczne. To właśnie ten kierunek poszukiwań ma największe umocowanie w praktyce, a o jego przydatności i kolo-

Pośród zastosowań praktycznych szczególne miejsce zajmują ułatwienia na polu teoretycznych, a więc ułatwienia w uprawianiu nauk. Chodzi tu o takie funkcje komputerów, które przyspieszają obliczenia, ułatwiają prezentację wyników (graficzną, dźwiękową itp.) czy wreszcie umożliwiają *symulacje*. Przeprowadzanie symulacji jest cenne zarówno dla nauk technicznych – jako pewien sposób wstępnego weryfikowania rozwiązań prototypowych (np. nowego silnika samochodowego), jak i dla nauk teoretycznych – jako metoda sprawdzania spójności i poprawności różnych modeli teoretycznych (np. psychologicznych modeli pamięci). O możliwości przeprowadzania symulacji przesądzają dwie cechy komputerów, które nie przysługiwały maszynom je poprzedzającym: możliwość kodowania w ich pamięci dowolnych stanów i procesów oraz błyskawiczna szybkość⁶.

Drugi rodzaj narzędzi, za które uznaje się często systemy informatyczne, to *narzędzia wyjaśniające*. Określenie to oddaje przekonanie niektórych, że czynności poznawcze ludzi przypominają w istotny sposób operacje wykonywane przez układy sztuczne, a budowa takich układów odpowiada jakoś strukturom poznawczym ludzi. Innymi słowy, wierzy się w to, że systemy informatyczne mogą stanowić adekwatny lub przynajmniej najlepszy z dotychczasowych model ludzkiego umysłu. Co tę wiarę wspiera? Przede wszystkim pewne analogie funkcjonalne między mózgiem a maszynami do przetwarzania informacji (zwróciliśmy na nie uwagę w ustępie 2), a ponadto ich niezwykła skuteczność w podejmowaniu zadań zarezerwowanych dotychczas dla ludzi⁷.

Myśląc o systemach informatycznych w kategoriach narzędzi wyjaśniających, wkraczamy na grunt o wiele bardziej grząski niż poprzednio. Wszak wnioskowanie z podobieństwa osiągniętych wyników o podobieństwie warunkujących je struktur czy zasad działania nie musi być uprawnione. Na przykład to, że maszyna rozpoznaje pewne kształty (choćaby litery) równie skutecznie jak człowiek, nie musi znaczyć, że jej system pojęciowy (podstawa kategoryzacji) jest zorganizowany analogicznie do ludzkiego. Aby zwrócić uwagę na wynikające tu trudności interpretacyjne, przeprowadźmy następujące rozumowanie dotyczące uczenia się.

⁶ Należy pamiętać, że większość systemów informatycznych – zwłaszcza tych, które wykonują zadania najambitniejsze – stanowi wytwór bardzo wielu specjalistów z wielu dziedzin. Za ich pośrednictwem może działać wprawdzie pojedynczy człowiek, ale wykorzystując je, ma świadomość obcowania z wiedzą wielokrotnie przewyższającą jego zdolności pamięciowe i operacyjne (między innymi – rzadko kiedy jeden użytkownik jest w stanie wykorzystać wszystkie możliwości użytkowanego systemu). Stąd, zdaniem rzeczników referowanego stanowiska, bierze się zapewne złudzenie wyższości maszyny nad człowiekiem i wiara w możliwość skonstruowania systemów autonomicznych.

⁷ Korzenie omawianego stylu myślenia o systemach informatycznych jako narzędziach wyjaśniających zdają się tkwić głęboko w ludzkiej kulturze. Ludzie zawsze mieli do wyboru dwie drogi uzyskiwania samowiedzy: albo samoobserwację i autorefleksję, albo analizę własnych wytworów. Ta druga droga może wydawać się okreźna, ale zawsze była bardzo owocna. Do szeroko pojętych wytworów, które stawały się wtórnie inspiracją refleksji humanistycznej, należą: logika, różnego rodzaju teorie fizyczne, maszyny.

Wyobraźmy sobie maszynę przyszłości, która rozwiązuje problemy wymagające w przypadku ludzi zdolności do uczenia się, lecz nie wykazuje innych cech towarzyszących uczeniu się, na przykład świadomości⁸. Czy zasady działania takiej maszyny można uznać za wyjaśnienie procesu uczenia się ludzi? Być może, ale z wieloma zastrzeżeniami. Wobec ubogiej (lub niejednolicie interpretowanej) wiedzy humanistów pozostanie kilka możliwości. Na przykład:

1. Maszyna wyjaśnia proces uczenia się ludzi. Wyjaśnienie to pokazuje, że owe inne cechy (takie jak świadomość) nie są konieczne do rozwiązywania problemów wymagających uczenia się, a człowiekowi przysługują one niejako „nadmiarowo”.
2. Maszyna wyjaśnia proces uczenia się w oderwaniu od innych. Mogą istnieć inne, bardzo ważne dla człowieka, a niezidentyfikowane przez humanistów, problemy, które wymagają połączenia zdolności do uczenia się z innymi.
3. Maszyna nic nie wyjaśnia. Istnieje wiele strategii rozwiązywania problemów wymagających uczenia się, a maszyna realizuje jedną z nich – równorzędną wobec strategii właściwych człowiekowi.

Przyjmowanie założenia o istnieniu istotnych analogii na styku umysł–system informatyczny może prowadzić do kolejnego poglądu na temat służebnej (tj. narzędziowej) roli informatyki. Nazwijmy go postrzeganiem systemów sztucznych jako *narzędzi treningowych*. Wiadomo wszak, że chcąc zrozumieć cokolwiek, nie da się uciec od ćwiczeń praktycznych. Na przykład, aby zrozumieć jakąś teorię fizyczną, trzeba przeprowadzać eksperymenty, aby pojąć istotę twierdzeń matematycznych, trzeba rozwiązywać odpowiednio dobrane zadania. Fakty te znają od dawna dydaktycy, a wyrazem ich znawstwa są chociażby programy nauczania na studiach wyższych, gdzie wykładom (teoria) towarzyszą zawsze ćwiczenia (praktyka). A zatem, czy o komputerach, programach komputerowych, algorytmach itp. nie można myśleć podobnie? Czy próbując konstruować „inteligentne” systemy informatyczne (a więc wnioskujące, rozpoznające mowę czy uczące się), nie pogłębiamy – na zasadzie treningu – własnej wiedzy o czynnościach poznawczych? Czy nie jest to prawdą tym bardziej, im bardziej naszym wysiłkom konstruktor-skim towarzyszy zamiar stworzenia maszyn na swój obraz i podobieństwo?

3. Maksymalizm oczekiwań

Powyższymi pytaniami zakończyliśmy przegląd intencji, jakie mogą kryć się za utożsamieniem: „system informatyczny to narzędzie”. Przejdźmy teraz na przeciwny biegun interpretacyjny i zbadajmy pogląd, który da się wyrazić następują-

⁸ Świadomość możemy traktować jako reprezentanta różnych cech „wewnętrznych”, związanych ze świadomymi przeżyciami człowieka, które należy oddzielić od cech „zewnętrznych”, przejawiających się w interakcji ze środowiskiem.

co: „Badania informatyczne są pożyteczne, bo stanowią punkt wyjścia do skonstruowania maszyn autonomicznych, maszyn będących równorzędnymi partnerami człowieka”. Innymi słowy: „Są pożyteczne, bo wiedą ku realizacji autonomicznych czynności poznawczych, a nie ku modelowaniu czy udoskonalaniu czynności poznawczych ludzi”. Pogląd taki można opatrzyć etykietą *maksymalizacji oczekiwań*. Nie znajduje on aż takiego oparcia w praktyce jak poprzedni, niemniej chcąc go uzasadnić, można powołać się na fakt nieustannego postępu w „umaszynawianiu” różnych funkcji poznawczych. Wszak komputery coraz lepiej zapamiętują, rozpoznają, komunikują się z ludźmi, uczą się, a ponadto funkcje te, implementowane dotychczas osobno, z coraz większym powodzeniem integruje się w pojedynczych systemach.

Maksymalizm oczekiwań jest bardzo nośny psychologicznie, dostarcza bowiem silnej motywacji do badań. Swoją siłę w tym względzie zawdzięcza śmiało postawionemu celowi, który dodatkowo znajduje swoje odbicie w starych jak świat mitach o istotach potężniejszych i jednocześnie zależnych od ludzi. Mimo tego wszystkiego ktoś, kto chciałby go zaakceptować, musi się liczyć z trudnościami natury zasadniczej. Rozważmy je kolejno.

Przede wszystkim, brak powszechnej zgody co do tego, jakie miałyby być owe stawiane maszynom *maksymalne wymagania*. Czy miałyby chodzić o zamiar „umaszynowania” samego intelektu, jak utrzymuje większość specjalistów w dziedzinie sztucznej inteligencji, czy należałoby dążyć do jakiejś realizacji świadomości, czy może trzeba by uwzględnić uczucia i emocje?

Nieco podobny problem stwarza *futurystyczne zabarwienie* maksymalizmu oczekiwań. Nie wiadomo przecież, jakie systemy do przetwarzania informacji zostaną zaprojektowane w przyszłości i które z nich okażą się najbardziej skuteczne. O tym, że monopolu w tym względzie być nie może, przekonuje krótka, lecz pouczająca historia informatyki. Są w niej obecne i maszyny cyfrowe, i maszyny analogowe, i sztuczne sieci neuronowe, i systemy selekcyjne. Nie wiadomo zatem, pod adresem jakich rozwiązań kierować swoje maksymalne oczekiwania. Od problemu można oczywiście uciec, twierdząc, że interesują nas wszelkie możliwe systemy do przetwarzania informacji – również takie, których zasad działania nie potrafimy jeszcze sobie wyobrazić. Niemniej, takie postawienie sprawy niezwykle utrudnia dyskusję z oponentami, a tym samym hamuje rozwój poglądów osoby wyznającej maksymalizm oczekiwań. Wzorowy przykład debaty o dobrze określonym przedmiocie stanowi dyskusja o możliwościach komputerów cyfrowych, w szczególności o szansach zrealizowania za ich pomocą funkcji świadomości⁹.

⁹ Debata ta jest owocna, ponieważ na jej wstępie przyjmuje się, że komputer cyfrowy to maszyna realizująca tradycyjne algorytmy, a więc dająca się opisać teoretycznie za pomocą formalizmu tzw. maszyny Turinga.

Jeśli chodzi o możliwość zrealizowania za pomocą tego typu urządzeń funkcji świadomości, to spotyka się dość szeroką paletę poglądów na ten temat. Zgodnie z poglądem skrajnie optymistycz-

Nawet po uściśleniu faktycznego zakresu oczekiwań i wskazaniu rodzaju systemów, pod adresem których je formułujemy, wynika zasadniczy problem *kryterium*. Mianowicie, na jakiej podstawie moglibyśmy w ogóle wnosić, że osiągnęliśmy ambitny cel skonstruowania systemu spełniającego nasze maksymalne oczekiwania. Czy zasadny byłby tu jakiś czysto „zewnętrzny” test nierozróżnialności maszyn i ludzi, czy też należałoby wziąć pod uwagę czysto humanistyczną wiedzę o przebiegu procesów poznawczych i sprawdzać ją z tym, co wiemy – jako pomysłodawcy i konstruktorzy – o zasadach działania maszyn. Kwestiom tym poświęcimy niżej nieco więcej miejsca.

W latach 50. XX wieku A. Turing, wybitny angielski matematyk o zacięciu filozoficznym, sformułował ideę bardzo prostego testu na inteligencję maszyn, tzw. *testu nierozróżnialności*. zilustrował go następującym obrazem. Wyobraźmy sobie *quiz*, polegający na zadawaniu pytań dwóm graczom, spośród których jeden to zamaskowany komputer. Pytania mają taką postać, że wystarczy odpowiadać na nie TAK lub NIE. Jeśli na podstawie udzielanych odpowiedzi nie jesteśmy w stanie odróżnić maszyny od człowieka, winniśmy uznać ją za inteligentną.

Chociaż pierwotna strategia Turinga dotyczyła wyłącznie badania inteligencji (pojmowanej dodatkowo jako zdolność do rozwiązywania problemów), dziś możemy uznać ją za przydatną do stwierdzania wszelkich cech charakterystycznych ludzkiego umysłu. Inteligencję, wrażliwość i inne dyspozycje ludzi porównujemy także w drodze specjalnych testów, a w życiu codziennym chętnie stosujemy maksymę „Sądź ludzi po ich uczynkach, a nie myślach”. Z punktu widzenia filozofii strategia ta stanowi wyraz wiary w zasadność operacjonizmu. Operacjoniści wierzą bowiem, że wszelkie „wewnętrzne” i prywatne cechy ludzi – takie jak świadomość, odczuwanie czy wola – daje się sprawdzać wyłącznie poprzez obserwacje ich zachowań zewnętrznych. Wyznają zatem hasło: „Myśleć znaczy odpowiednio działać”¹⁰. W odniesieniu do maszyn testowanych metodą Turinga działanie to przejawiałoby się w krótkich i jednoznacznych odpowiedziach na postawione pytania.

Opisany test na inteligencję maszyn (a w naszym ujęciu również na inne dyspozycje) razi niektórych zbytnią prostotą. Dlatego też doczekał się wielu modyfikacji, które w zamierzeniu autorów mają odpowiadać temu, co naprawdę rozumiemy przez „myślenie”. Stanisław Lem – niekwestionowany autorytet w dzie-

nym (znanym pod angielską nazwą „strong artificial intelligence”, co tłumaczy się często jako „silna wersja sztucznej inteligencji”) dla otrzymania świadomości wystarczy odpowiednio zaawansowany algorytm, a podłoże materialne, w którym zostanie on uruchomiony (układ elektroniczny, mechaniczny czy biologiczny), nie ma większego znaczenia. Według stanowiska skrajnie pesymistycznego, które wywodzi się z kartezjańskiego dualizmu duszy i ciała, dla zaistnienia świadomości nie wystarczy żaden algorytm ani żadne podłoże materialne. Pomiedzy zarysowanymi skrajnościami występuje wiele poglądów pośrednich. Ich popularny opis – Penrose, R., *Nowy umysł cesarza*, PWN, Warszawa 2000 – daje W. Marciszewski w cytowanej pracy *Sztuczna inteligencja*.

¹⁰ Penrose, R., *Nowy umysł cesarza*, dz. cyt., s. 24.

dzinie futurologii – zaproponował następującą zmianę. Niech maszyna nie odpowiada na pytania, niech wysłucha czyjejś historii, a następnie spróbuje opowiedzieć ją własnymi słowami. Dopiero owe „własne słowa” pozwolą nam przekonać się, czy maszyna myśli równie twórczo i indywidualistycznie jak człowiek¹¹. Z kolei współczesny polski logik i filozof Witold Marciszewski zwrócił uwagę na konieczność sprawdzania niewiedzy maszyn. Według niego testowany komputer winien nie tylko odpowiadać na pytania, ale również domagać się ich większej precyzji, uzasadnień, wyjaśnień itp. A zatem winien dążyć do rozjaśnienia swojej niewiedzy. Decyzję o nierozróżnialności człowieka i maszyny moglibyśmy podjąć dopiero na podstawie analizy jej pytań. Dlaczego jest to takie ważne? Otóż aby komputer mógł zadać sensowne pytanie, musi czegoś nie wiedzieć, a ponadto ów brak wiedzy musi go niepokoić i prowokować do szukania wyjaśnień. Innymi słowy: testowany komputer powinien mieć własne problemy i wolę ich rozwiązywania¹².

Zauważmy, że wzmiankowane przed chwilą rozszerzenia testu Turinga odchodzą od idei skrajnego operacjonizmu w stronę bliższej analizy rozumienia przekazywanych maszynie tekstów. U Lema przejawia się to wprost – jako nakaz sparafrazowania pewnej historii, u Marciszewskiego pośrednio – poprzez odniesienie się pytaniem do treści, które nie zostały z jakichkolwiek powodów, do końca zrozumiane.

Tym samym tropem podąża inna wielka postać debaty na temat przyszłości systemów informatycznych, tym razem zagorzały przeciwnik operacjonizmu wszelkiej maści, J. Searle. Około dwudziestu lat temu filozof ten przedstawił rozumowanie, nazywane dzisiaj *argumentacją chińskiego pokoju*¹³. Pokazuje ono dobitnie różnice między sposobem rozwiązywania problemów przez człowieka i zaprogramowany komputer. Wyobraźmy sobie – proponuje Searle w jednej z wersji swojej argumentacji – zamknięty pokój, w którym pracuje osoba dokonująca przekładów tekstów angielskich na chińskie. Osoba ta nie zna języka chińskiego. Tłumaczeń dokonuje w oparciu o zestaw precyzyjnych instrukcji typu: „gdy zobaczysz taką a taką literę, takie a takie słowo lub taką a taką konstrukcję gramatyczną języka angielskiego, zastąp ją takim a takim odpowiednikiem chińskim”. Tłumacz wcielający w życie wzmiankowane instrukcje postępuje dokładnie tak jak komputer. Gdy na dostarczonej kartce rozpozna odpowiedni ciąg symboli, sięga do katalogu instrukcji, wypisuje na ich podstawie odpowiednie symbole chińskie, po czym zwraca tłumaczenie klientowi. Wszystko odbywa się automatycznie i mechanicznie, bez żadnego wnikania w znaczenia zastępowanych symboli. Na klientach, nieświadomych tajników jego warsztatu, tłumacz sprawia wrażenie oso-

¹¹ Lem, S., *Bomba megabitowa*, Wydawnictwo Literackie, Kraków 1999, s. 195–199.

¹² Marciszewski, W., *Sztuczna inteligencja*, Znak, Warszawa 1998, s. 78–81.

¹³ Searle J.R., *Umysł na nowo odkryty*, PIW, Warszawa 1999, s. 71, 72.

by rozumiejącej chiński wyśmienicie. Tymczasem my wiemy, iż mimo sprawnego przekładania tekstów język chiński to dla niego coś zupełnie niezrozumiałego (mówiąc żartobliwie, zupełna „chińszczyzna”). Przykład wymyślony przez Searle’a daje przeciwnikom operacjonizmu naprawdę mocne oparcie. Widać żeń jak na dłoni, że czym innym jest zachowywać się inteligentnie, a czym innym wykazywać cechy, które w przypadku ludzi warunkują takie zachowanie. Przecież nawet gdyby „lokator chińskiego pokoju” osiągał swoje wyniki na innej, mniej bezmyślnej drodze, i tak nie mielibyśmy pewności, że jego operacje przypominają (a więc wyjaśniają) postępowanie człowieka.

Przewrotność przykładu Searle’a wyraża się w tym, że to człowiek, którego inteligencję chcielibyśmy wyjaśnić za pomocą komputerowych analogii, zniża się do poziomu maszyny i to on imituje jej operacje. Naśladując maszynę, staje się bezmyślnym wykonawcą rozkazów, a tym samym zatracą wszelkie cechy rzeczywistej inteligencji. Skryta za złośliwością intencja Searle’a wydaje się jasna. Należy odwrócić rolę: zamiast z podobieństwa zewnętrznych zachowań wnioskować, iż komputery wykazują cechy typowo ludzkie, znane nam działania komputerów trzeba zawsze oceniać z punktu widzenia tego, co wiemy o ludziach, a więc z perspektywy wiedzy czysto humanistycznej.

Na początku niniejszego ustępu poruszyliśmy kwestię nieprzewidywalnej przyszłości systemów informatycznych. Na jego zakończenie wypada poświęcić nieco miejsca *rozwojowi samego człowieka*. Sprawa to niezwykle ważna, a musi się uporać z nią każdy, kto na serio stawia przed maszynami maksymalne wymagania. Chodzi o niewątpliwy fakt sprzężenia zwrotnego między nami a naszymi wytworami. Przecież im więcej będą potrafiły komputery, tym więcej będziemy potrafili my. Pozostanie to prawdą zarówno wtedy, gdy będziemy używać ich jako narzędzi, jak i wtedy, gdy będziemy wchodzić z nimi w równorzędne relacje. Innymi słowy: wzrost „mocy poznawczej” systemów informatycznych będzie pociągał za sobą wzrost „mocy poznawczej” ludzi. Czy zatem można mówić z sensem o programie ostatecznej, maszynowej realizacji funkcji poznawczych człowieka?

5. Uwagi końcowe

Pytanie zadane na końcu ostatniego akapitu stanowi dobrą okazję do poczynienia podsumowującej tekst deklaracji światopoglądowej. Deklaracja ta będzie jednocześnie zawierać szkieletową odpowiedź na postawione pytanie. Otóż wydaje nam się, że warunkiem koniecznym rozwiązania problemu rozwoju jest wszczepienie w systemy informatyczne mechanizmów uczenia się, a więc dynamicznej adaptacji do zmian otoczenia.

Na czym jednak miałyby polegać sugerowane mechanizmy adaptacyjne? Obserwacja ludzi podpowiada, że powinny to być zarówno zmiany indywidualne, jak i gatunkowe, a więc w przypadku maszyny zarówno zmiany programu nadzo-

rującego pracę konkretnej maszyny, jak i zmiany długoterminowe, rozciągnięte na wiele kolejnych populacji maszyn. Procesy pierwszego rodzaju przypominałyby uczenie się pojedynczych ludzi, procesy drugiego rodzaju miałyby za swój wzór ewolucję. Zmiany, o których mowa (można o nich myśleć jako o zmianach struktury wewnętrznej i zasad działania pewnych programów komputerowych), byłyby też niezdeterminowane. Wyobrażamy to sobie w ten sposób, że ich kierunek byłby określony przez ogólne cele maszyn (np. dążyć do minimalizacji zużycia energii), lecz ich szczegółowy przebieg ustalałby się w dużej mierze przypadkowo. Zauważmy, że maszyny samodzielnie modyfikujące własne działanie – tak w skali jednostkowej, jak i populacyjnej – byłyby wolne od najpoważniejszego ograniczenia współczesnych systemów informatycznych. Mianowicie, kierujące nimi programy nie byłyby dokładnymi przepisami działań we wszystkich możliwych okolicznościach (rozpoznanych wcześniej przez zespoły projektantów, programistów i testerów), lecz składałyby się ze zbioru ogólnych celów, do których maszyna winna dążyć, i odpowiednich procedur adaptacji.

Informatycy stawiający sobie za cel skonstruowanie opisywanych systemów są w dość komfortowej sytuacji. Mogą bowiem połączyć własną znajomość tajników programowania i budowy systemów informatycznych z ideami i narzędziami wypracowanymi w ramach innych dziedzin. Wymieńmy dla przykładu kilka z nich. Logika daje im do dyspozycji okazały arsenał różnorodnych technik wnioskowania – gdy samo wnioskowanie połączy się z weryfikowaniem przydatności i zapisem jego wyników, uzyskuje się bardzo ciekawą metodę uczenia się. Neurobiologia wzbogaca ich wiedzę o samoorganizacji struktur neuronowych, tworzących między innymi ludzki mózg. Teoria ewolucji i genetyka pokazują, na czym polega rozwój i adaptacja w skali całych populacji.

Pozostaje dodać, że informatycy w rzeczywistości sięgają do tak odległych źródeł inspiracji. Na przykład dla trzech wymienionych wyżej dziedzin można wskazać trzy ogólne modele przetwarzania informacji. Są to: tradycyjne algorytmy (ze szczególnym uwzględnieniem algorytmów pisanych w językach logikopodobnych), sieci neuronowe i algorytmy genetyczne. Modele te wykorzystuje się od jakiegoś czasu w bardzo nowatorskiej dziedzinie informatyki, określanej zbiorczą nazwą „uczenie się maszyn” (ang. „*Machine learning*”). To właśnie jej dynamiczny rozwój motywuje autora niniejszej pracy do wiary, że odpowiedź na pytanie postawione w jej tytule jest pozytywna.

Bibliografia

- Bolc, L., Zaremba, J., *Wprowadzenie do uczenia się maszyn*, Akademicka Oficyna Wydawnicza RM, Warszawa 1992.
- Bolter, J.D., *Człowiek Turinga. Kultura Zachodu w wieku komputera*, PWN, Warszawa 1998.

- Chwedeńczuk, B. (red.), *Filozofia umysłu*, PWN, Warszawa 1999.
- Cichosz, P., *Systemy uczące się*, WNT, Warszawa 2000.
- Hetmański, M., *Umysł a maszyny. Krytyka obliczeniowej teorii umysłu*, Lublin 2000.
- Lem, S., *Bomba megabitowa*, Wydawnictwo Literackie, Kraków 1999.
- Marciszewski, W., *Sztuczna inteligencja*, Znak, Warszawa 1998.
- Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Programs*, Springer Verlag, 1992.
- Mitchell, T., *Machine Learning*, McGraw-Hill, 1997.
- Penrose, R., *Nowy umysł cesarza*, PWN, Warszawa 2000.
- Russell, S., Norvig, P., *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1994.
- Searle, J.R., *Umysł na nowo odkryty*, PIW, Warszawa 1999.
- Tadeusiewicz, R., *Sieci neuronowe*, Akademicka Oficyna Wydawnicza RM, Warszawa 1993.

Will Computer Scientists Be Able to Create Autonomous Machines?

The question in the title points to a deep dissimilarity between man—an entity of the highest known level of autonomy, and a machine—an object built by man. The machines now existing have not developed beyond the status of object-tools. Their multifarious functions—practical, explanatory or training—have all been designed by man, are controlled and utilized by man, too. Is it possible therefore that information systems in the future will gain some autonomy? A rational discussion of this question meets with three problems (1) it is not clear what kind of units would be considered in the discussion, (2) it is not clear how such machines should be distinguished from people, (3) it is clear, however, that humans interact with the machines they have build, and in the course of that interaction they become more advanced and capable. The author argues that the last difficulty is essential. To overcome the barrier separating machines from humans, it would be necessary to construct machines that can interact with their environment and are capable of self-perfection. Some developments in computer sciences look promising and seem to support the possibility that such machines will be created some day. These developments have already offered some valuable insights to psychology, neurobiology and theory of evolution.