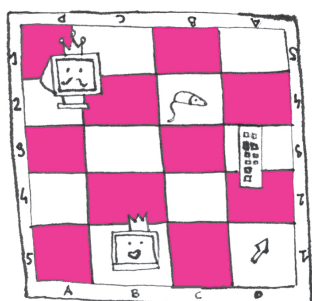


Jest to skrót artykułu opublikowanego w *Przeglądzie Filozoficznym* 22 (2013), nr 2, 167–184.

Termin „sztuczna inteligencja” jest dwuznaczny, bo używa się go zarówno do określania dziedziny badawczej, jak i jej potencjalnego produktu. Aby tej dwuznaczności uniknąć, będziemy używać tego terminu tylko w pierwszym znaczeniu, rezerwując dla drugiego pojęcia określenie „maszyna myśląca”.



Zwycięstwo *Deep Blue* nad szachistą Kasparowem w 1997 roku było możliwe między innymi dlatego, że uniknięto tam prostego przeszukiwania przestrzeni możliwych ruchów.



## Gdzie jesteś, HAL?

Jarek GRYZ\*

Umysł to komputer. Teza ta, która nawet w tak niedoprecyzowanej formie jest dla wielu z nas nie do przyjęcia, legła u podstaw sztucznej inteligencji, jednej z najbardziej fascynujących i kontrowersyjnych dziedzin nauki zapoczątkowanych w ubiegłym stuleciu. Była to dziedzina, która bodaj jako pierwsza wyodrębniła się z informatyki jako osobna poddziedzina, ale miała ambicje daleko poza informatykę wykraczające. Przyciągnęła ona najtęższe umysły matematyki, ekonomii, psychologii i filozofii, ale skonsumowała też ogromne (prawdopodobnie największe w informatyce) fundusze badawcze. Jej zadufanie i bombastyczne obietnice z pierwszych lat istnienia z czasem ustąpiły przesadnej wręcz skromności i chęci zdegradowania dziedziny przez nazywanie jej racjonalnością obliczeniową zamiast sztuczną inteligencją.

**Początki.** Pierwszy amerykański komputer, ENIAC, został uruchomiony w 1945 roku. Jego zastosowania były czysto wojskowe i dotyczyły symulacji wybuchu i potencjalnych zniszczeń spowodowanych przez projektowaną wówczas bombę wodorową. W powszechnym mniemaniu komputer to było po prostu szybkie liczydło. Warto o tym pamiętać, bo trzeba było nie lada geniuszu, by wyobrazić sobie inne zastosowania dla owego „liczydła”. Geniuszem tym wykazał się Herbert Simon, który pracował wówczas nad komputerową symulacją obrony powietrznej. Tak oto narodził się paradygmat komputera jako maszyny do przetwarzania informacji: komputer operuje na symbolach odnoszących się do obiektów istniejących realnie. Stąd był już tylko krok do funkcjonalnego zidentyfikowania komputera i ludzkiego umysłu.

Niemal od pierwszej chwili, kiedy odkryto, że komputer może przetwarzać dowolne symbole (a nie tylko liczby), usiłowano stworzyć program, który pokazałby możliwości maszyny w tej dziedzinie. Oczywiście zastosowaniem były szachy: gra, której reguły łatwo jest opisać w języku symbolicznym, a w której – jak się początkowo wydawało – szybkość przetwarzania symboli (pozycji na szachownicy i ich „wartości”) miała kluczowe znaczenie dla pokazania przewagi maszyny nad człowiekiem. Wnet jednak okazało się, że „bezmyślne” przeszukiwanie wszystkich sekwencji ruchów na szachownicy dla znalezienia takiej, która prowadziłaby nieodzwrotnie do wygranej, jest nierealne. Claude Shannon oszacował liczbę możliwych posunięć na  $10^{120}$ , co znaczy, że przy weryfikacji jednej sekwencji w ciągu jednej milionowej sekundy pierwszy ruch na szachownicy wykonany byłby po  $10^{95}$  latach. Herbert Simon i współpracujący z nim wówczas Allen Newell nie zamierzali oczywiście tak długo czekać. Zmienili więc dziedzinę i postanowili napisać program, który dowodziłby twierdzeń logiki. Choć złożoność obliczeniowa takiego programu jest dużo mniejsza niż gry w szachy, to i w tym przypadku konieczne było sformułowanie reguł i metod efektywnego poszukiwania dowodu, a nie proste weryfikowanie, czy dowolnie wygenerowany ciąg znaków spełnia wymogi dowodu w sensie logicznym. Program, który w ten sposób powstał, *Logic Theorist*, był więc w pewnym sensie „kreatywny”, bo generował dowody, których jego autorzy się nie spodziewali. Innymi słowy, choć program spełniał wymagania postawione przez jego autorów, jego zachowania nie można było (przynajmniej w łatwy sposób) przewidzieć. Sukces programu był spektakularny: dowiódł on 38 z 52 twierdzeń drugiego rozdziału *Principia Mathematica* Russella i Whiteheada. Co więcej, dowód twierdzenia 2.85 sformułowany przez *Logic Theorist* okazał się bardziej elegancki niż ten z *Principiów*. Russell był pod wielkim wrażeniem owego sukcesu, ale już „The Journal of Symbolic Logic” odmówił publikacji dowodu, którego autorem była maszyna. Te wczesne, niewątpliwe sukcesy sztucznej inteligencji skłoniły wielu ówczesnych naukowców zarówno do przeceniania własnych osiągnięć, jak i do stawiania hurraoptymistycznych prognoz na przyszłość.

Oczywiście, badacze sztucznej inteligencji zdawali sobie sprawę, że z faktu, iż pewne elementy ludzkiego myślenia mogą być imitowane czy realizowane przez komputer, nie wynika, że dotyczy to wszystkich elementów. Potrzebny był więc argument pokazujący, co dla myślenia jest istotne, a co jest tylko przypadkowe. Przypadkowe byłoby więc, na przykład, fizyczne umiejscowienie procesów

\*York University, Toronto, Kanada



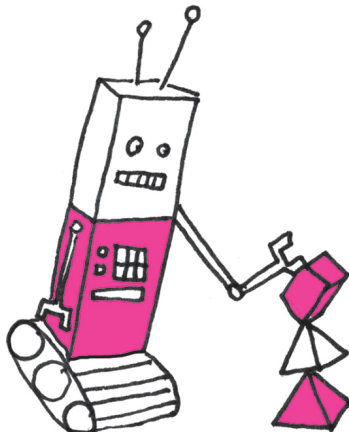
### Rozwiązanie zadania M 1439.

Niech  $f(x) = ax^2 + bx + c$ . Mamy

$5 \mid f(0) = c$ . Następnie

$5 \mid f(1) - f(-1) = 2b$ , więc  $5 \mid b$ . Ponadto,

$5 \mid f(1) = a + b + c$ , zatem również  $5 \mid a$ .



### Projekt SHRDLU

Celem SHRDLU było zbudowanie robota, który potrafiłby porozumiewać się z człowiekiem, rozumować i przeprowadzać proste działania. I tak SHRDLU dysponował chwytakiem, za pomocą którego mógł manipulować rozrzuconymi na stole klockami o różnych kolorach i kształtach. Komunikacja między takim „robotem” a człowiekiem odbywała się za pomocą klawiatury oraz monitora, na którym pokazany był stół z klockami i odpowiedzi robota. Kiedy wydawano robotowi polecenie ułożenia klocków w jakiejś konfiguracji, SHRDLU wymyślał plan działania i wirtualnie go wykonywał. SHRDLU potrafił zrozumieć język potoczny i do pewnego stopnia rozumiał swoje motywy. Oto przykład interakcji między SHRDLU a człowiekiem:

- Podnieś czerwony klocek.
  - OK. [SHRDLU podnosi klocek]
  - Znajdź klocek wyższy niż ten, który trzymasz, i włóż go do pudełka.
  - Zakładam, że przez „go” rozumiesz klocek, który trzymam. [SHRDLU wkłada właściwy klocek do pudełka]
  - Co jest w pudełku?
  - Niebieska piramida i niebieski klocek.
  - Czy stół może podnosić klocki?
  - Nie.
  - Czy można postawić piramidę na piramidzie?
  - Nie wiem.
  - Postaw piramidę na piramidzie.
  - Nie mogę. [SHRDLU próbuje, ale mu się nie udaje]
  - Niebieska piramida jest moja.
  - Rozumiem.
  - Czy jest w pudełku coś, co jest moje?
  - Tak, niebieska piramida.
- Na pierwszy rzut oka dialog ten sprawia wrażenie, jakby rzeczywiście rozmawiało się z maszyną myślącą, która rozumie, co się do niej mówi. Czytelnik Eloquentny wymyśli jednak zapewne bez trudu sekwencję pytań prowadzącą do udzielenia przez maszynę odpowiedzi absurdalnie bezsensownych.

myślowych w mózgu; zastąpienie jednego czy wręcz wszystkich neuronów przez elementy mechaniczne (takie jak obwody elektryczne) o tej samej funkcji nie powinno mieć wpływu na procesy myślowe.

Co zatem stanowi o istocie myślenia i co pozwala nam ignorować konkretną realizację procesów myślowych? Newell i Simon sformułowali to w postaci słynnej hipotezy systemu symboli jako warunku koniecznego i wystarczającego dla inteligentnego działania. Manipulacja czy też obliczanie przy użyciu tych symboli to właśnie myślenie.

Warto zauważyć, że w powyższej wersji hipotezy (tzw. wersji silnej) ludzki umysł musi być takim systemem symboli (w przeciwnym razie należałoby mu bowiem odmówić możliwości myślenia). W takiej wersji zbudowanie maszyny myślącej przy użyciu komputera jest, oczywiście, teoretycznie wykonalne, z tej prostej przyczyny, że ludzki umysł to po prostu komputer. Wersja silna hipotezy ma jeszcze jedną istotną konsekwencję, a mianowicie, że badając działania komputera (czy, ściślej rzecz biorąc, zainstalowanego w nim programu), możemy dowiedzieć się czegoś nowego na temat działania mózgu. Na założeniu tym ufundowana została kognitywistyka. Zauważmy na koniec, że dla uzasadnienia celów sztucznej inteligencji wystarczy słaba wersja hipotezy, mianowicie, że manipulacja symbolami jest wystarczająca, ale niekonieczna dla myślenia.

**Rozwój.** Lata sześćdziesiąte i siedemdziesiąte XX wieku to złoty wiek sztucznej inteligencji. Gwałtownie wzrosła ilość badaczy zajmujących się tą dziedziną, a także funduszy przeznaczonych na badania. Większość rozwiązań algorytmicznych i systemowych, które weszły do kanonu dziedziny, pochodzi właśnie z tamtych czasów. Co więcej, sukcesy sztucznej inteligencji przestały już być tylko wewnętrzną sprawą naukowców. Media zapowiadały cywilizacyjną rewolucję, która lada dzień dokonać się miała za sprawą budowanych właśnie maszyn myślących. Ale za tą efektowną fasadą pojawiły się też pierwsze pęknięcia. Zdawano sobie już wcześniej sprawę z ułomności i ograniczeń budowanych systemów, ale dopiero teraz zaczęto wyraźnie rozpoznawać problemy. Jak się później okazało, wielu z nich nie dawało się rozwiązać za pomocą usprawnień technologicznych, bo wynikały one z dokonanych wcześniej założeń metodologicznych czy też filozoficznych, na których opierała się sztuczna inteligencja.

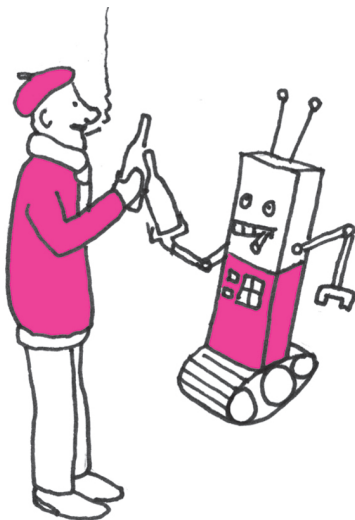
Na czym polegał problem? Otóż wszystkie programy stworzone do tej pory w sztucznej inteligencji stosowały się do tzw. mikroświatów, a więc ściśle zdefiniowanych i dokładnie opisanych wycinków rzeczywistości. Tak więc *Logic Theorist* dowodził tylko twierdzeń logiki, a SHRDLU (patrz obok) manipulował tylko klockami na stole. Wszelkie próby rozciągnięcia zastosowań tych programów na szersze dziedziny, zdefiniowane nie tak ściśle jak logika czy gra w warcaby, kończyły się fiaskiem. Bariery były dwojakiego rodzaju: pierwsza dotyczyła złożoności obliczeniowej, druga – wiedzy potocznej. Warto od razu podkreślić, że żadnej z tych barier nie udaje się pokonać za pomocą doskonalszych algorytmów, bardziej skomplikowanych programów czy szybszych komputerów. Zmiany wymaga – jak się wydaje – sam paradygmat sztucznej inteligencji. Przyjrzyjmy się zatem bliżej każdej z wymienionych barier.

**Problemy.** Zadania, jakie wykonać ma program komputerowy, są zaimplementowane w postaci algorytmów. Efektywność programu zależy zarówno od fizycznych własności komputera (wielkości pamięci, szybkości CPU itd.), jak i ilości operacji, jakie wykonać musi algorytm. Ten drugi czynnik zależy z kolei od ilości danych na wejściu (łatwiej znaleźć maksimum z 10 liczb niż z 1000 liczb) oraz od skomplikowania samego algorytmu (łatwiej znaleźć maksimum z 10 liczb, niż je posortować). Złożoność obliczeniowa to właśnie miara tego skomplikowania, a definiuje się ją po prostu jako funkcję  $f(n)$ , gdzie  $n$  jest ilością danych na wejściu. Algorytmy o złożoności wykładniczej są w praktyce nieobliczalne; mogą one być wykorzystywane wyłącznie do rozwiązywania problemów o małej skali (tzn. dla niewielkiego  $n$ ). Takie właśnie problemy rozwiązywano w mikroświatach: na przykład SHRDLU manipulował tylko kilkoma klockami. Wykorzystanie tych samych



### Rozwiązanie zadania F 868.

Przed umieszczeniem na kuli ładunku przyciąga się ona z cząstką w wyniku wyindukowania ładunku na jej powierzchni. Niech siła ta wynosi  $F_{\text{IND}}$ . Jeżeli ładujemy kulę kolejno ładunkiem  $q, 2q, 3q$ , to pojawia się dodatkowa siła odpychania odpowiednio  $F, 2F, 3F$ . Znajdując wypadkową siłę działającą pomiędzy kulą i cząstką, w każdym z tych przypadków, dostajemy  $F_1 = F_{\text{IND}} - F$ ,  $F_2 = F_{\text{IND}} - 2F$  i  $F_3 = F_{\text{IND}} - 3F$ . Stąd ostatecznie  $F_3 = 2F_2 - F_1$ . Zauważmy, że siła ta może być siłą przyciągania albo odpychania.



programów, czy też, ściślej mówiąc, tych samych algorytmów, do rozwiązywania realistycznych problemów jest po prostu niemożliwe. Wydawać by się mogło, że tego dylematu można jednak uniknąć: skoro barierą są algorytmy o złożoności wykładniczej, dlaczego nie zastosować innych, szybszych algorytmów dla rozwiązywania tych samych problemów? Otóż, jednym z największych osiągnięć teorii informatyki ostatniego wieku była obserwacja, że szybszych algorytmów dla większości problemów sztucznej inteligencji najprawdopodobniej nie ma! Udało się do dziś zidentyfikować kilkaset problemów, które określa się mianem NP-zupełnych, a które mają dwie wspólne im własności. Po pierwsze, dla żadnego z tych problemów nie udało się do tej pory znaleźć rozwiązania o mniejszej niż wykładnicza złożoności. Po drugie, znalezienie szybkiego, a więc wielomianowego rozwiązania dla jednego z nich rozwiązuje w ten sam sposób je wszystkie. Dla sztucznej inteligencji był to wynik szczególnie dotkliwy, bo większość problemów, które usiłowano w sztucznej inteligencji rozwiązać, należy właśnie do klasy NP-zupełnych lub jeszcze trudniejszych. Dla wielu krytyków był to argument za tym, że sztuczna inteligencja oparta na rozwiązaniach algorytmicznych, czyli obliczeniowych, jest niemożliwa.

Drugi poważny problem, który napotkano w sztucznej inteligencji, dotyczył opisu wiedzy potocznej. Jeśli maszyna myśląca ma wchodzić w interakcje z ludźmi i funkcjonować w ich środowisku, musi – choćby częściowo – podzielać ich obraz świata. Wiedza potoczna, która ten obraz świata opisuje, musi być zatem w jakiś sposób reprezentowana w języku maszyny.

Prawdopodobnie nikt nie zdawał sobie sprawy, jak trudny to może być problem, aż do czasu publikacji artykułu, w którym McCarthy i Hayes zdefiniowali tzw. problem ramy. Autorów interesowało sformalizowanie myślenia zdroworozsądkowego na użytek planowania działań przez robota. Skonstruowany w tym celu tzw. rachunek sytuacji pozwalał opisywać rezultaty działań i przeprowadzać stosowne wnioski. Rozważmy następujący problem. Jeśli w sytuacji  $S$  obiekt  $x$  jest na obiekcie  $y$  i  $x$  jest pusty (tzn. nic na nim nie ma), to w rezultacie zdjęcia obiektu  $x$  z obiektu  $y$  w sytuacji  $S$  obiekt  $y$  również będzie pusty. A co z kolorem obiektu  $y$ ? Czy zmieni się on w wyniku zdjęcia z niego obiektu  $x$ ? Dla nas odpowiedź jest jasna: nie. Ale odpowiedź ta nie jest wcale oczywista z punktu widzenia opisywanej teorii. Sytuacja, która powstała w wyniku zdjęcia obiektu  $x$  z obiektu  $y$ , jest inną sytuacją niż sytuacja  $S$ . Dopóty, dopóki nie stwierdzimy *explicite*, że kolor obiektów nie zmienia się w wyniku ich przenoszenia, nie wiemy, czy obiekt zachował swój kolor. Problem wydaje się trywialny, ale bynajmniej nie ma trywialnego rozwiązania na gruncie formalizmu logicznego. Dodanie tzw. aksjomatów ramy, a więc twierdzeń typu „obiekt nie zmienia swego koloru w wyniku jego przenoszenia” jest nie do przyjęcia z trzech powodów. Po pierwsze, w warunkach realnego świata takich aksjomatów byłoby nieprzewidywalnie wiele, a mianowicie tyle, ile jest różnych par działanie-własność. Dodawanie nowych własności i nowych działań do opisu świata wymagałoby nieustannego dodawania takich aksjomatów. Po drugie, prawdziwość takich aksjomatów zależy od kontekstu. Jeśli jeden robot przenosi klocki, a drugi je maluje, to powyższy aksjomat jest fałszywy. Wreszcie, jeśli aksjomaty ramy miałyby opisywać sposób myślenia człowieka w sytuacjach takich, jak opisana powyżej, to na pewno opisują ją fałszywie.

Kolejny problem jest dość oczywisty i dotyczy trudności, a właściwie niemożliwości wyliczenia wszystkich skutków naszych działań. Problem ten nazywamy problemem rozgałęzionych efektów. Okazuje się jednak, że nie tylko skutków, ale również warunków naszych działań nie jesteśmy w stanie wyliczyć. Klasyczny przykład to próba uruchomienia samochodu. Spodziewamy się, że wystarczy do tego przekręcić kluczyk w stacyjce. Okazuje się, że to nieprawda: akumulator musi być naładowany, w baku musi być benzyna, rura wydechowa nie może być zatkana kartoflem, nikt w nocy nie ukradł nam silnika itp. Trudno oczekiwać, że bylibyśmy w stanie podać wszystkie warunki niezbędne do podjęcia jakiegokolwiek działania. Które z nich powinniśmy zatem wyspecyfikować dla zdefiniowania działań dostępnych dla robota? To tzw. problem uszczegółowienia warunków.



### Rozwiązanie zadania M 1440.

Odp. Jest  $2^{(m-1)(n-1)}$  takich macierzy.

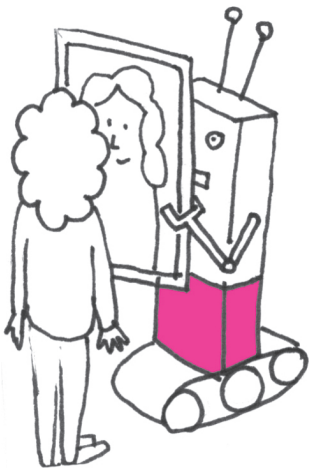
Zauważmy, że podmacierz złożona z pierwszych  $m - 1$  wierszy (o parzystej liczbie jedynek) wyznacza jednoznacznie ostatni wiersz macierzy (macierz należy uzupełnić, wstawiając 1 do kolumn z nieparzystą liczbą 1 w pierwszych  $m - 1$  wierszach i 0 w pozostałych kolumnach). Zauważmy, że tak dodany wiersz zawiera parzystą liczbę jedynek – w przeciwnym przypadku łączna liczba jedynek w pierwszych  $m - 1$  wierszach byłaby nieparzysta.

Pojedynczy wiersz o parzystej liczbie jedynek można wybrać na  $2^{n-1}$  sposobów – jest to liczba podzbiorów zbioru  $n$ -elementowego o parzystej liczbie elementów. Zatem  $m - 1$  wierszy można wybrać na  $(2^{n-1})^{m-1} = 2^{(n-1)(m-1)}$  sposobów, co jest równe liczbie interesujących nas macierzy.

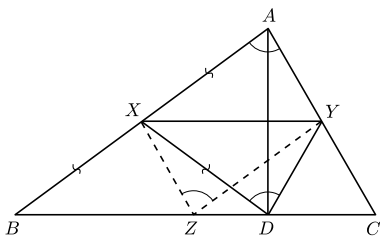




Wydaje się, że trudno byłoby dziś znaleźć badaczy, którzy zgadzaliby się z założeniem biologicznym. Natomiast bez założenia ontologicznego opis świata za pomocą języka byłby chyba niemożliwy.



**Rozwiązanie zadania M 1438.**  
Niech  $Z$  będzie środkiem boku  $BC$ .



Ponieważ trójkąty  $AXY$  i  $ZYX$  są przystające, to  $\sphericalangle XZY = \sphericalangle XAY = \sphericalangle XDY$ . W takim razie punkty  $X, Y, D, Z$  leżą na jednym okręgu. Stąd trapez  $XYZD$  jest równoramienne, w szczególności  $XD = ZY = \frac{1}{2}AB = XB = XA$ . Punkty  $A, B, D$  leżą więc na okręgu o środku  $X$  i promieniu  $\frac{1}{2}AB$ . Innymi słowy, punkt  $X$  jest środkiem okręgu opisanego na trójkącie  $ABD$  i jednocześnie jest środkiem boku  $AB$  tego trójkąta. Ten trójkąt musi zatem mieć kąt prosty przy wierzchołku  $D$ .

**Rozpad.** Najsurowsza krytyka sztucznej inteligencji nadeszła ze strony filozofii, głównie za sprawą Johna Searle'a i Huberta Lederera Dreyfusa. Słynny argument „chińskiego pokoju” Searle'a zdawał się pokazywać, że niemożliwe jest, aby program komputerowy cokolwiek „rozumiał” czy też miał „świadomość”. Tym samym program budowy maszyny *myślącej* przy użyciu programów komputerowych jest skazany na niepowodzenie. Tymczasem krytyka Dreyfusa skierowana jest wprost przeciwko sztucznej inteligencji i podważa jej pozostałe, być może nawet istotniejsze, założenia. Tezy, przeciwko którym występuje Dreyfus, to, po pierwsze, *założenie biologiczne*: mózg to mechanizm manipulujący symbolami, podobnie jak komputer, po drugie, *założenie psychologiczne*: umysł to mechanizm manipulujący symbolami, podobnie jak komputer, po trzecie, *założenie epistemologiczne*: inteligentne zachowanie można sformalizować i tym samym skopiować przy użyciu maszyny, i wreszcie *założenie ontologiczne*: świat składa się z niezależnych, nieciągłych faktów.

Komercyjne systemy eksperckie budowane w latach osiemdziesiątych XX wieku nie okazały się wielkimi sukcesami. Zdobycie, sformalizowanie, a następnie właściwe wykorzystanie specjalistycznej wiedzy w jakiejś dziedzinie było problemem nie tylko technicznym, ale dotyczyło znowu problemu wiedzy potocznej. Po raz kolejny obietnice dokonywane w sztucznej inteligencji, tym razem w wydaniu komercyjnym, nie sprawdziły się. Dla większości badaczy stało się oczywiste, że postęp, jaki dokonał się w ciągu 40 lat przy projekcie zbudowania maszyny myślącej, a więc takiej, która przeszłaby test Turinga, był znikomy. Ogrom pracy, funduszy i ludzkich talentów zaangażowanych w tę dziedzinę był niewspółmierny do osiągniętych efektów. Co więcej, wielu badaczy miało poczucie, że droga, którą obrano, prowadzi donikąd. Z drugiej strony, nawet najzagorzalsi krytycy programu sztucznej inteligencji musieli przyznać, że w niemal każdej z jej poddziedzin dokonania były znaczące. Nic zatem dziwnego, że to właśnie w tych poddziedzinach koncentruje się dziś wysiłek badaczy. Począwszy od lat dziewięćdziesiątych, ze sztucznej inteligencji wyodrębniła się robotyka, przetwarzanie języka naturalnego, teoria decyzji i wiele innych. Każda z tych dziedzin organizuje swoje własne konferencje i wydaje pisma naukowe, w niewielkim tylko stopniu wchodząc w interakcje z innymi dziedzinami. Tak oto, dla większości badaczy, podniosły cel zbudowania maszyny myślącej stał się celem ulotnym czy wręcz nieistotnym.

Czy jest zatem coś, co spaja dziedzinę nazywaną wciąż jeszcze sztuczną inteligencją? Jaki cel przyświeca tym wszystkim, którzy usiłują zbudować systemy percepcji, podejmowania decyzji, wnioskowania czy uczenia się? Wydaje się, że wyróżnić można co najmniej trzy stanowiska w tej sprawie. Pierwszą – stanowiącą dziś mniejszość – postawę reprezentują ci, którzy nadal uważają, że celem sztucznej inteligencji jest zbudowanie maszyny imitującej zachowanie (a więc i myślenie) człowieka. Druga postawa apeluje o odrzucenie ludzkiego zachowania i myślenia jako wzorca dla sztucznej inteligencji i uważają, że najważniejsze jest budowanie maszyn, które przewyższają ludzkie zdolności umysłowe w jakiś użyteczny sposób. Dla uzasadnienia skuteczności tej metody badacze przywołują przykład aeronautyki, która odniosła sukces dopiero wówczas, gdy przestała w konstrukcji samolotów imitować ptaki. Trzecia wreszcie postawa, najbardziej, jak się wydaje, powszechna, to całkowite porzucenie idei budowy maszyny myślącej. Zamiast tego proponuje skoncentrowanie na budowie urządzeń i systemów, które maksymalizują zadaną miarę wydajności.

Druga z opisanych wyżej postaw jest najbliższa programowi kognitywistyki. Badamy procesy poznawcze – ludzkie, zwierzęce czy maszynowe – nie po to, by zbudować imitację człowieka, ale po to, by go zrozumieć. Wykorzystujemy w tym celu dane z każdej nauki, która nam na ten temat coś może powiedzieć: psychologii, neurofizjologii, lingwistyki czy właśnie sztucznej inteligencji w ostatnim rozumieniu. Projektowanie robotów nie musi być już celem samym w sobie, ale środkiem do zrozumienia zachowań i procesów myślowych ludzi i zwierząt. Wielu badaczy obarcza test Turinga odpowiedzialnością za wyznaczenie nieosiągalnego, niepotrzebnego i niewiele znaczącego wzorca dla sztucznej inteligencji.

Wydaje się więc, że HAL 9000 na długo pozostanie tylko filmową fantazją.