Tytuł: **Maszyna Turinga a umysł ludzki**

Autor: Marek Hetmański / hetman@ramzes.umcs.lublin.pl

Źródło: http://www.kognitywistyka.net / mjkasperski@kognitywistyka.net

Data publikacji: 09 VIII 2002

Termin 'maszyna Turinga' odnosi się do teoretycznego projektu maszyny matematycznej sformułowanego w latach trzydziestych przez Alana M. Turinga. Jest on szeroko wykorzystywany i dyskutowany także poza matematyką w psychologii poznawczej, teoriach sztucznej inteligencji, jest podstawą tzw. komputacyjnej koncepcji umysłu. Zamiarem artykułu jest analiza teoretycznej treści maszyny Turinga (pewnych jej ograniczeń) oraz ocena użyteczności tego pojęcia w psychologicznych i filozoficznych koncepcjach ludzkiego umysłu. Teza jest następująca – maszyna Turinga nie może być właściwym (poprawnym) modelem umysłu i działania ludzkiego; może być niemniej użyteczna (w ograniczonym zakresie) w analizie niektórych czynności poznawczych człowieka. Kwestią otwartą jest to, jakie inne jeszcze modele mogą być pomocne w opisie poszczególnych rodzajów myślenia i działania człowieka; czy inne rodzaje maszyn (np. homeostat cybernetyczny, sieci neuronowe, uniwersalny komputer kwantowy, czy inne rodzaje maszyn analogowych) mogą symulować całość (może tylko jakiś aspekt) działań człowieka?

1. Problem rozstrzygalności w matematyce. Rozstrzygalność a algorytmizacja

Algebraizacja logiki przeprowadzona przez Boole'a, rozwinięta potem przez wielu innych autorów, doprowadziła w latach dwudziestych i trzydziestych obecnego stulecia do badań nad podstawami matematyki. W ich ramach postawiono szereg ważkich kwestii, również takie, które mają teoriopoznawcze znaczenie i są dyskutowane poza matematyką. Jedną z nich jest problem **rozstrzygalności**. Jest to problem takiej własności aksjomatycznych systemów, która polega na tym, że w większości przypadków można podać warunki ich obliczalności przez zastosowanie funkcji rekurencyjnych (funkcji obliczalnych). Funkcje takie w skończonej liczbie kroków podają warunki rozstrzygnięcia tego, czy dane twierdzenie jest elementem systemu, czy metoda tego rozstrzygnięcia jest efektywna. Efektywna metoda jest **algorytmem**.

Zagadnienie rozstrzygalności podjął Turing w swojej koncepcji maszyny matematycznej. Chcąc podać warunki obliczalności, efektywnego rozwiązania danego zadania matematycznego sformułował abstrakcyjne, czysto teoretyczne pojęcie automatu, który samoczynnie wykonuje pewne proste operacje na symbolach w celu podania rozwiązania tego zadania. Automat ten wykonuje swoje operacje analogicznie do działań każdego rachmistrza wykonującego proste czynności rachunkowe, jak zapisywanie danych liczbowych i pośrednich wyników, posługiwanie się określonymi symbolami i regułami, dochodzenie do rozwiązania zadania. Wstępnym zamiarem Turinga było wykazanie, że wszelkie efektywnie rozwiązywalne (obliczalne, algorytmizowalne) zadanie matematyczne może być wykonane przez taki automat.

Podstawowym sformułowaniem maszyny matematycznej do podawania warunków rozstrzygalności zagadnień matematycznych jest artykuł Turinga z 1936/37 roku pt. *On Computable Numbers with an Application to the Entscheidungsproblem*. Wystąpienie Turinga zbiegło się w czasie i było równorzędne co do wartości z osiągnięciami A. Churcha, E. Posta i S. Kleene'a; przypisuje się mu jednakże największą rangę ze względu na najwyraźniej sformułowane założenie o możliwości mechanizacji obliczeń, jak również samo używanie słowa 'maszyna' (por. Gandy 1988). W swoim artykule Turing odniósł się do zagadnienia (funkcjonowało ono wówczas jeszcze pod niemiecką nazwą) sformułowanego przez Dawida Hilberta. Wyrażało się ono w pytaniu czy istnieje pewna ogólna, mechaniczna procedura rozstrzygania ogólnej klasy poprawnie sformułowanych problemów matematycznych? Inaczej mówiąc, czy dla takich zagadnień istnieje algorytm podający warunki rozwiązania zagadnienia? Oczekiwanie Hilberta, co do możliwości podania procedur algorytmicznych dla dowolnego zagadnienia matematycznego (w tym wyrażał się jego program formalistycznej interpretacji matematyki) zostało, skrótowo mówiąc, przez Turinga (podobnie jak przez innych) zasadniczo podważone. Oryginalnym i ważnym jego wkładem w to zadanie jest podanie istotnych warunków, ale również ograniczeń, procedur mechanicznych w odniesieniu do bardzo abstrakcyjnie i szeroko zdefiniowanej klasy maszyn matematycznych. Dzięki niezwykle sugestywnemu pomysłowi Turinga owocnie zaczęto rozważać nie tylko istotne kwestie metamatematyczne, ale również konstruować cyfrowe maszyny liczące.

2. Maszyna Turinga – podstawowe założenia

Maszyna Turinga jest tworem wyłącznie teoretycznym, swoistą grą umysłową, konstruktem, który miał służyć jego autorowi rozwiązaniu ważnego metamatematycznego problemu. Określenie 'maszyna Turinga' wprowadził do użycia po raz pierwszy A. Church w recenzji z artykułu Turinga. Turinga nie interesowało na samym początku rozważań, to czy można skonstruować fizyczną maszynę, która dokonałaby algorytmicznych obliczeń. Dopiero potem (w trakcie wojny i po niej, gdy brał udział w pracach nad łamaniem szyfrów maszyn kodujących) zagadnienie to stało się dla niego praktyczną kwestią. W artykule z 1936/37 roku Turing za punkt wyjścia przyjął konstrukcję abstrakcyjnego rachmistrza, który dokonuje obliczeń z użyciem bardzo elementarnych przedmiotów, jak kartki z pokratkowanego zeszytu do rachunków, na których zapisuje proste znaki na potencjalnie nieskończonej taśmie. Postawił przy tym fundamentalne pytanie: „Jakie są możliwe procesy, które mogą być wykonane podczas obliczania?" Miał przy tym na myśli dosłowne czynności wykonywane przez rachmistrza, które mogą też być wykonane przez zaprojektowaną maszynę; użycie zwrotu 'mechaniczne wykonanie' znaczyło w tym kontekście tyle, co „możliwe do wykonania przez maszynę". Turing przyjął, że czynności mechanicznego obliczania są ograniczone, podobnie jak ograniczone są zmysłowe zdolności każdego rachmistrza (obejmuje wzrokiem tylko pewną część kratek na taśmie) oraz jego umiejętności umysłowe (zapamiętuje pewną tylko ilość reguł postępowania podczas obliczania); pod tym względem istotne matematyczne pojęcie ma za przesłankę pewne psychologiczne założenie.

R. Gandy[1] referując podstawowe założenia Turinga wprowadził na określenie ludzkiego rachmistrza termin ‘komputer’ (w przeciwieństwie do ‘komputera’ oznaczającego fizyczną realizację maszyny matematycznej), którego działanie charakteryzuje się:

- liczbą różnych symboli zapisywanych w kratkach;
- liczbą przyległych kratek, których treść komputer może rozpatrzyć (Turing przyjął, że dla rachmistrza czytającego kratki w układzie linearnym liczba ta jest mniejsza niż 15);
- możliwością zmiany w danym kroku komputera treści tylko w jednej kratce;
- ilością „stanów umysłu” komputera; jego stan umysłu wraz z treścią przeglądanej kratki wyznacza działanie komputera i następny stan jego umysłu; komputer wykonuje zawsze ustalony, skończony zbiór instrukcji.

Ze względu na wymóg poglądowości maszynę Turinga przedstawia się również w literaturze tematu (także tej popularnej) w mniej lub bardziej fizycznym kształcie graficznych schematów (już bez psychologicznych implikacji, które czynił sam Turing). Na ich postać mają przy tym (co jest zrozumiałe) wpływ elementy z późniejszych technicznych, konstruktorskich projektów komputera według tzw. architektury von Neumanna. (Wzajemny wpływ obu matematyków w rozwoju maszyn liczących jest zresztą do dzisiaj tematem badań i analiz[2].

Na treść pojęcia maszyny Turinga składają się zatem następujące elementy, których nie można jednak uważać w dosłownym znaczeniu za części maszyny[3]:

- jednostka centralna (kontrolna), która określa dowolną ilość trybów pracy maszyny;
- skończony zbiór nie zmieniających się w czasie pracy maszyny reguł postępowania, dowolnie jednak wymienialny;
- sekwencja klatek w swobodnie przesuwanej taśmie, na której maszyna zapisuje/wymazuje znaki;
- rejestr stanów maszyny (od stanu wyjściowego do stanu końcowego), w ramach którego realizuje się zawsze określony algorytm przypisany maszynie w danym zadaniu. Te elementy są konieczne, aby móc efektywnie podejść do zagadnienia rozstrzygalności.

Teoretyczne składowe maszyny Turinga przedstawia się również (taki jest wymóg poglądowości) za pomocą pewnej liczby (jej wielkość zależy od stopnia dokładności opisu) fizycznych elementów, głównie w następujących postaciach:

- czytnika;
- taśmy o nieograniczonej długości z wyróżnionymi kratkami, na których może znajdować się znak, który może być zmieniany w trakcie pracy maszyny; kratka może zawierać bądź tylko jeden (z co najmniej dwóch wyróżnionych i wykluczających się znaków), bądź być pusta; ilość znaków stosowanych przez maszynę Turinga może być dowolna, lecz zawsze skończona, przy czym najczęściej stosowanym systemem znakowym jest układ binarny: 1 i 0, dzięki któremu maszyna Turinga ma faktycznie do czynienia z trzema możliwościami, może też wykonywać operacje również w stosunku do kratki pustej.

---

[1] R. Gandy, *The Confluence of Ideas in 1936*, s. 81

[2] Por. M. Davies, *Mathematical Logic and the Origin of Modern Computers*, ss. 165-169.

[3] Por. R. Ligonniere, *Prehistoria i historia komputerów*, Ossolineum, Wrocław 1992, ss. 205-214.

Istotą działania maszyny Turinga jest etapowe, sekwencyjne wykonywanie kolejnych podstawowych działań. Każde jej działanie określone jest przez tryb narzucony przez jednostkę kontrolną (stan maszyny) oraz znak odczytywany z taśmy. W każdym stadium swojego działania maszyna Turinga może zatem wykonać na poszczególnych poziomach następujące czynności:

- na poziomie jednostki kontrolnej przejść z jednego trybu pracy w drugi lub utrzymać aktualny;
- na poziomie urządzenia zapisu/odczytu w stosunku do danej kratki po jej odczytaniu maszyna może: (1) zapisać znak, jeśli kratka jest pusta, (2) wykasować znak i zastąpić go innym znakiem lub pozostawić wolne miejsce, (3) nie zmieniać nic; po czym może przejść o jedną kratkę w lewo lub w prawo, bądź też pozostać w miejscu.

Czynności te układają się w listę (siatkę) poleceń, które maszyna ma wykonać. Sprowadzają się one do działań w stosunku do danej kratki (z trzema powyższymi możliwościami), zachowania lub zmiany stanu maszyny przed kolejną operacją wobec kratki oraz przesunięcia taśmy o jedną kratkę w prawo lub lewo. Ilość poleceń (instrukcji) jest zawsze skończona, układają się one w listą, która w całości determinuje pracę maszyny. Lista instrukcji danej maszyny, zapisana na taśmie, może być również listą poleceń dla pewnej innej maszyny Turinga; jest to ważna cecha maszyny zaprojektowanej przez Turinga (o czym później) decydująca o jej uniwersalności.

Wszystkie działania wykonywane przez maszynę Turinga dyktowane są określonym z góry **programem,** na który składają się (z funkcjonalnego punktu widzenia) kombinacje trybów pracy jednostki kontrolnej (określają one jaki rejestr poleceń ma być zastosowany) oraz odczytywanie określonego znaku. Z mechanicznego punktu widzenia działanie maszyny jest sekwencją dyskretnych przejść z jednego stanu w drugi i wykonywaniem operacji na znakach zapisanych na taśmie. Maszyna Turinga pracuje przywołując jedną tylko na raz regułę ze skończonego ich zbioru. Odpowiednio do niej operuje znakiem na taśmie i odwołuje się do reguły kolejnej aż do momentu, gdy przywołana reguła nie zatrzyma maszyny. Maszyna zatem "wie" dwie rzeczy: którą regułę wykonuje i jakim znakiem z taśmy operuje. Reguła i znak determinują jednoznacznie jej sekwencyjne działanie.

Powyższą, wyłącznie formalną, charakterystykę maszyny matematycznej można uzupełnić charakterystyką z punktu widzenia teorii informacji. Znaki w kratkach taśmy można bowiem zinterpretować jako informację (dane) a operacje na nich (zamiana znaku jednego na inny) jako przetwarzanie informacji. Przy założeniu możliwości dowolnie bogatego słownika znaków oraz dowolnie zmienianego programu można powiedzieć, że zasadniczo **maszyna Turinga działa wobec dowolnej informacji**; sposób kodowania informacji (zapis danych) jest obojętny. W tym poszerzeniu (zbieżnym z cybernetyką) koncepcji maszyny Turinga leży źródło wielu prób używania jej jako modelu nie tylko matematycznego automatu czy cyfrowej maszyny liczącej (komputera), ale również umysłu człowieka.

3. Uniwersalność i ograniczenia maszyny Turinga

Każda maszyna Turinga ma swoją określoną **moc**, czyli **zdolność rozwiązywania złożonych zadań**. Jest ona funkcją możliwych do przybrania przez nią stanów oraz bogactwa słownika, czyli przyjętych znaków. Maszyny Turinga można łączyć (teoretycznie) w dowolne układy. Dwie maszyny o takiej samej strukturze budowy (ilości stanów i bogactwie słownika) mają taką samą moc. Możliwe jest połączenie kilku uzupełniających się maszyn o różnej strukturze, lecz przeznaczonych do realizacji jednego określonego zadania, do wykonywania bardziej złożonych zadań.

Jest wiele różnych wariantów prostej maszyny Turinga, które można konstruować poprzez poszerzanie jej zasadniczych elementów – taśmy i znaków. W miejsce jednej taśmy można wprowadzić wiele taśm, równolegle odczytywanych i zmienianych przez urządzenie odczytu/zapisu. Jednowymiarową taśmę można także zastąpić dwuwymiarową płaszczyzną a nawet trójwymiarową przestrzenią, co daje o wiele większe możliwości zapisu i odczytu danych. W pewnym sensie całe otoczenie maszyny Turinga może być potraktowane jako taśma z zapisanymi znakami. Niemniej jednak ta możliwość pozornie tylko poszerza zdolności maszyny Turinga, gdyż w istocie operowanie przez nią poszerzoną i rozbudowaną ilością danych i tak sprowadza się do operowania w danym momencie znakiem z jednej kratki jednowymiarowej taśmy; płaszczyznowe czy przestrzenne ujęcie danych do przetworzenia tylko rozbudowuje i wydłuża czas operacji. Możliwości teoretyczne (moc obliczeniowa) maszyny Turinga nie zależą od jej parametrów „technicznych", lecz od zasady działania. Istotną jest w każdym przypadku **nieskończoność** taśmy (płaszczyzny czy przestrzeni) z zapisanymi danymi.

Wariantowość dotyczy tak samo drugiego elementu jej budowy – znaków zapisywanych na kratkach taśmy. Może on być również dowolnie poszerzany. Tradycyjnie stosowany zapis binarny (0 i 1) jest o tyle wygodny, że odpowiada ważnej własności fizycznej realizacji (późniejszej w stosunku do projektu) maszyny Turinga w postaci cyfrowego komputera, w którym impulsy zmiennego prądu elektrycznego (włączenie lub wyłączenie przełącznika w komputerze lampowym lub niskie i wysokie napięcie impulsu w tranzystorze komputerów nowych generacji) są fizycznym podłożem zapisu dwójkowego. Ma on jednak czysto konwencjonalne, do pewnego stopnia przypadkowe znaczenie. Można bowiem zastosować dowolnie bogatszy, zawsze jednak skończony, zbiór znaków o innej podstawie (dziesiętnej, ósemkowej itp.), który da większe możliwości operowania znakami, lecz mimo tego nie zmieni to istoty działania maszyny Turinga. Rozszerzony system dwójkowy stosowany w komputerach cyfrowych pozwala zapisywać nie tylko dowolną liczbę naturalną, lecz również liczby ujemne, ułamki. Modyfikacje systemu kodowania pozwalają również na binarny zapis nie tylko liczb, ale również wzorów matematycznych – algebraicznych, trygonometrycznych, dzięki czemu odpowiednio skonstruowane maszyny Turinga mogą wykonywać operacje na wzorach i regułach.

Turing rozważył możliwość poszerzenia mocy maszyny matematycznej. W stosunku do zwykłych maszyn Turinga wykonujących proste zadania można zbudować jedną wyróżnioną maszynę. Należy listę (siatkę) poleceń, instrukcji dla dowolnej maszyny Turinga zakodować w postaci ciągu symboli 0 i 1 oraz zapisać na taśmie. Taśmę tą następnie trzeba wykorzystać jako początkową część danych dla pewnej szczególnej maszyny – nazwanej przez Turinga – **uniwersalną maszyną**, która w stosunku do pozostałych danych z taśmy działa podobnie, jak działałaby maszyna zwykła. Skrótowo mówiąc, uniwersalna maszyna przejmuje jako część swojego programu program maszyny zwykłej. Uniwersalna maszyna Turinga potrafi zatem udawać każdą inną dowolną maszynę Turinga, może ją symulować. Wszystkie współczesne komputery są uniwersalnymi maszynami Turinga.

Właśnie ta teoretyczna cecha maszyn matematycznych daje niektórym teoretykom sztucznej inteligencji podstawą do postawienia pytania: czy umysł (niektórzy pytają także o mózg, jako szczególny rodzaj maszyny cyfrowo-analogowej) jest również uniwersalną maszyną Turinga? Pytanie to formułuje się również inaczej: czy maszyna Turinga symuluje działanie umysłu?

Maszyna matematyczna ma jednak ważne ograniczenia formalne, które każą ostrożnie traktować jej analogie z ludzkim umysłem; niemniej porównanie ograniczeń obu układów (nawet odmiennej natury) może być ważne i pouczające. Paradoksalnie w stosunku do przyjętych założeń o rozstrzygalności (obliczalności) oraz roli algorytmów w mechanizacji dowodzenia matematycznego, Turing wykazał, że nie ma uniwersalnego algorytmu, który można by zastosować do wszystkich problemów matematycznych, do wszystkich maszyn Turinga. Chociaż każda maszyna matematyczna dowodząc danego (poprawnie sformułowanego) twierdzenia arytmetycznego realizuje dany algorytm, to jednak nie istnieje algorytm, który dowiódłby, że maszyna ta wykona swoje obliczenia.

4. Maszyna Turinga a twierdzenie Gödla

Istotnym założeniem formułowanym od początku w badaniach nad mechanizacją dowodzenia matematycznego była teza mówiąca, iż każde zadanie teoretyczne, które da się opisać precyzyjnie może zostać zakodowane arytmetycznie i być wykonane w zaprojektowanej maszynie. Sens matematyczny tej tezy łączy się jednak z wieloma trudnościami i jest w pewnej części ograniczony. Jednak w praktyce, w próbach automatyzacji (komputeryzacji) obliczeń program ten jest niemniej częściowo realizowany. Komputerowe dowodzenie prawdziwości pewnych twierdzeń (np. problemu czterech barw rozwiązanego przez komputer w 1977 r.) nie tylko pozwala na rozwiązywanie starych matematycznych problemów, ale również zwraca uwagę na ważne cechy procesu myślenia twórczego (zasadniczo niealgorytmizowalnego) w naukach formalnych.

Mechanizacja dowodu matematycznego łączy się z ważnym zagadnieniem teoretycznym, o którym mówi twierdzenie Kurta Gödla. Zostało ono sformułowane w 1931 roku w trakcie dyskusji nad programem formalizmu matematycznego Hilberta. Stwierdza się w nim, że **każdy niesprzeczny system arytmetyki jest niezupełny,** tj. istnieje takie prawdziwe zdanie tego systemu o liczbach naturalnych, którego prawdziwości nie można udowodnić w ramach tego systemu. Niedowodliwość dowolnego twierdzenia dedukcyjnego systemu za pomocą jego własnych środków obala zasadniczo formalistyczny program logiki i matematyki. Powstaje wówczas pytanie czy godzi to w możliwość pełnej mechanizacji i algorytmizacji dowodów matematycznych? Sens twierdzenie Gödla dotyczy bez wątpienia maszyny Turinga (z czego Turing zdawać sobie sprawę). W odniesieniu do operacji wykonywanych przez maszynę matematyczną sens twierdzenia Gödla jest jednoznaczny – procedura obliczeniowa nie obejmuje wszystkich dowodów poddanych mechanizacji. Nie wszystkie dowody matematyczne są algorytmizowalne, prawdę w matematyce można uzyskiwać także na innej drodze. Ten wniosek natury metamatematycznej (i epistemologicznej) wywołał ożywioną dyskusję wśród matematyków.

Gödel był wyraźnie przekonany, że prawda jest dowodliwa także na innej drodze. Mając wprawdzie na uwadze twierdzenia Churcha i Turinga, pisał jednak:

> (...) na podstawie dotychczas udowodnionych twierdzeń nie można wykluczyć możliwości, że istnieje maszyna do dowodzenia twierdzeń, w rzeczywistości

> równoważna matematycznej intuicji (i nie wykluczone, że nawet uda się ją odkryć empirycznie), natomiast nie można udowodnić, iż dana maszyna jest równoważna matematycznej intuicji i generuje tylko prawdziwe twierdzenia z zakresu skończonej teorii liczb. [R. Penrose, *Makroświat, mikroświat i ludzki umysł*, s. 117].

Był zdania, że nie można wykluczyć, że matematycy posługują się poprawną procedurą algorytmiczną, której poprawności nie mogą jednak tą drogą udowodnić, że mogą także posługiwać się procedurami zasadniczo niemechanicznymi. Sens powyższej wypowiedzi Gödla jest istotny z formalnego i epistemologicznego punktu widzenia. W pierwszym przypadku, zasadniczo obala program formalizmu, gdyż wykazuje, że pojęcie dowodu matematycznego jako obliczalnego ciągu twierdzeń w obrębie zupełnego systemu aksjomatycznego jest nie do utrzymania, w drugim zaś pokazuje, że pojęcie prawdy danego twierdzenia nie może być definiowane czysto syntaktycznie (gramatycznie) jako przekształcanie wyrażeń. Pojęcie prawdy i dowodu wykraczają tym samym poza formalne granice systemu. R. Gandy zwraca uwagę na to, że Gödel (podobnie jak von Neumann) nie podał formalnego dowodu na rzecz nierozwiązywalności *Entscheidungsproblem* (chociaż zgadzał się z wnioskami Turinga), ponieważ był przede wszystkim zajęty – w opozycji do panującego wówczas klimatu intelektualnego – analizą niefinitystycznych pojęć i metod.

> A zainteresowanie niefinitystycznym wnioskowaniem nie jest niezbędne dla analiz obliczeń. Gödel podziwiał i akceptował analizy Turinga, nie jest jednak zaskakujące, że nie uczestniczył w nich. W rzeczywistości do samego końca swojego życia wierzył, że możemy być w stanie użyć niefinitystycznego wnioskowania w (niemechanicznych) obliczeniach. [R. Penrose, *Makroświat, mikroświat i ludzki umysł*, s. 69].

Podobne stanowisko zajmował Emil Post, którego badania i wnioski antycypowały w sporej części odkrycia Gödla, Churcha i Turinga. Analizując pojęcie systemu formalnego był przekonany, że w rozstrzyganiu jego spójności ma miejsce przewaga intuicji matematyka w stosunku do procedur mechanicznych. Pewne systemy można rozpoznać jako spójne na długo wcześniej niż wie się jak tego dowieść.

> Ustanowienie tezy nie jest sprawą matematycznego dowodu, lecz psychologicznej analizy umysłowych procesów zawartych w kombinatorycznych metematycznych procesach. (...) Czyni to matematyka kimś więcej niż rodzajem bystrej istoty, która może wykonać szybko to, co maszyna mogłaby zrobić w ostateczności. Widzimy, że maszyna nie mogłaby dać nigdy kompletnej logiki, w stosunku do zbudowanej maszyny możemy dowieść twierdzenia, którego ona nie może. [M. Davies, *Mathematical Logic and the Origin of Modern Computers*, ss. 415-417].

Chociaż nie jest wykluczone, jak sugeruje Gandy[4], że Post mówiąc 'maszyna' miał raczej na myśli maszynę w znaczeniu mechanicznego urządzenia a nie abstrakcyjną maszynę Turinga, to faktem pozostaje, że (podobnie jak Gödel, później inni) zakłada przewagę specyficznie ludzkiego czynnika (intuicji) nad wyłącznie mechanicznym, maszynowym. Znalazło to wyraz w jego rozróżnieniu między procedurami decyzyjnymi a procedurami wytwarzania poprawnych stwierdzeń.

Problem powyższy wciąż wywołuje wiele kontrowersji także poza matematyką i rodzi rozbieżne stanowiska. W literaturze filozofii umysłu (anglosaskiej tradycji analitycznej) przykładem rozważań nad powyższymi pytaniami była dyskusja rozpoczęta przez Johna Lucasa, która wywołała (w przeciągu dwóch dziesięcioleci) liczne komentarze i krytykę ze

[4] R. Gandy, *The Confluence of Ideas in 1936*, s. 95.

strony wielu autorów (np. D. Hofstadter, D. Dennett, H. Wang, D. Lewis). W podsumowaniu (po latach) całej dyskusji Lucas stwierdza, że dla wykazania tezy, iż ludzkie umysły nie są maszynami Turinga można skutecznie posłużyć się twierdzeniem Gödla, gdyż mówi ono nie tylko o prawdzie w systemach formalnych, lecz także o prawdzie mającej związek z umysłem; prawdy te funkcjonują jednak w obydwu układach (systemach) w odmienny sposób. Twierdzenie Gödla nie jest jednak w tej kwestii rozstrzygnięciem ostatecznym. Trzeba wyjść od następujących faktów: mechaniczne (algorytmiczne) dowodzenie w systemach formalnych nie jest tożsame z prawdziwością jego twierdzeń; prawda a dowód są w wielu przypadkach pojęciami rozłącznymi; przypisywanie przez człowieka atrybutu „prawdziwy” niektórym twierdzeniom może odbywać się na drodze niealgorytmicznej; finitystyczna interpretacja systemów formalnych stoi w pewnej opozycji z nieskończonością umysłowych zdolności człowieka (w tym doskonaleniem dowodów); ta nieskończoność poznawczych zdolności z kolei stoi w opozycji do skończoności życia (np. czasu na dowodzenie). Te różne cechy systemów formalnych i wiedzy (umysłu) człowieka wzajemnie się ograniczają i jednocześnie warunkują. Lucas stwierdza, że poszukiwanie modelu umysłu w maszynie całkowicie niesprzecznej jest nieuzasadnione nie tylko z powodu ograniczeń, jakie nakłada na nią twierdzenie Gödla, ale głównie dlatego, że człowiek dochodzi do prawdziwości znacznej klasy twierdzeń także drogą pozaformalną, uznaje swoją wiedzę za wartościową poprzez intuicję. Nawet jeśli dokonuje formalizacji sposobów dowodzenia, wartościowania, sądzenia, mówienia itp., to czynność ta nie jest w pełni kompletna, algorytmizowalna.

> Nie twierdzą – pisze Lucas – że Gödlowski argument nie może być sformalizowany, lecz to, że (jakąkolwiek formalizację przyjmiemy) istnieją inne argumenty, które są w wyraźny sposób wartościowe, chociaż nie obejmuje ich ta formalizacja. Musimy być zawsze gotowi rozeznać bez szczególnych trudności pewne stosowane reguły wnioskowania, lecz musimy również, jeśli mamy być racjonalni, poszerzyć zakres uznawanych za wartościowe wnioskować poza uprzednio ustanowione granice. Nie wyklucza to następnie ich formalizowania, lecz nie możemy zakładać, że każda formalizacja jest indukcyjnie kompletna. [J. Bobryk, *Akty świadomości i procesy poznawcze*, ss. 114-115].

Formalizacji można poddać wszystko, także samą formułę Gödlowską (niedowodliwą w systemie, chociaż prawdziwą), co tylko pozornie jest paradoksalne. Trzeba bowiem rozróżnić pomiędzy dowodliwością w systemie formalnym a nieformalną dowodliwością dostarczoną przez (formalizowany) Gödlowski argument.

Możliwy zatem do przyjęcia jest taki maszynowy model umysłu, w którym byłby on wprawdzie maszyną operującą rachunkiem zdań (np. wypowiedziami), lecz maszyna symulująca jego działanie musiałaby posiadać także instrukcje jak sprawdzać czy rewidować porządek aksjomatów (pewne z nich musiałyby być niezmienione), odwołując się przy tym do pozaformalnych racji. Lucas stwierdza, że ludzkie umysły, przy pewnej ogólnej interpretacji, są takimi maszynami. Nie są to jednak maszyny niesprzeczne, lecz raczej niespójne. Niespójność w „maszynerii” ludzkiego umysłu wyraża się szczególnie w aktach mowy, gdzie podmiot nie przystępuje do wypowiadania twierdzeń w jednolitym (niesprzecznym) słownictwie, a wręcz przeciwnie – formułuje wszelkie rodzaje nonsensów czy sprzeczności, komunikując je w słownictwie pełnym gaf językowych i niejednoznaczności, także w formie werbalnej i pozawerbalnej (gesty). Zdaniem Lucasa (argument ten podnosił w dyskusji również Dennett) ten fakt w stopniu o wiele większym niż epistemologiczna implikacja twierdzenia Gödla wskazuje na różnicę umysłu wobec maszyny Turinga.

Przekonanie, że możliwości poznawcze człowieka (np. dowodzenie czy używanie predykatu „prawdziwy") są innej natury niż skuteczność maszyny matematycznej w dowodzeniu niesprzeczności systemów formalnych nie jest powszechne, nie jest też bezdyskusyjne. Porównywanie efektywności człowieka wykonującego operację obliczania funkcji rekurencyjnych z efektywnością maszyny Turinga może prowadzić do różnych wniosków. Warta odnotowania jest uwaga H. Putnama, który zakłada, że nawet jeśli maszyna Turinga (T) rzeczywiście nie jest w stanie dowieść rozstrzygalności danego twierdzenia (U) systemu dedukcyjnego, to o jego prawdziwości przesądza się na innej drodze; postępuje zresztą tak maszyna, jak i człowiek.

> Jednakże T może równie dobrze dowieść tego samego, tj. że U jest dlań nierozstrzygalne i że jeżeli T jest niesprzeczna, to U jest 'prawdziwe' na mocy zaprogramowanej interpretacji. Zaś zdania U, którego T **nie może** udowodnić (przy założeniu jej niesprzeczności), i **ja** bynajmniej dowieść nie mogę (dopóki nie udowodnię, że T jest niesprzeczna, co w przypadku, gdy T jest bardzo skomplikowana, jest mało prawdopodobne)! [1961, s. 142].

Maszyna może zatem, podobnie jak człowiek, dowieść, że dla pewnego zdania nie jest w stanie podać dowodu, a także – jeżeli jej program jest niesprzeczny – że zdanie to jest jednak prawdziwe. Możliwości maszyn matematycznych w zakresie wykonywania operacji matematycznych nie są zatem mniejsze niż możliwości umysłu ludzkiego.

Podobnie sądzi M. Scriven, gdy pisze:

> Twierdzenie Gödla wskazuje na trudność, która nie jest większa w przypadku maszyny niż w przypadku nas samych. Można tylko stwierdzić, że matematyka byłaby łatwiejsza, gdyby formaliści mieli rację, i że wówczas zbudowanie mechanicznego matematyka byłoby rzeczą stosunkowo prostą. Jednakże tak nie jest. Natomiast rozpoznanie prawdziwości niedowodliwej formuły przez porównanie tego, co ona mówi z tym, co już znamy jako prawdziwe, jest dostępne w tym samym stopniu dla człowieka, jak i dla maszyny. [1961, s. 125].

Twierdzenie Gödla nie jest w myśl tej opinii argumentem ostatecznie zaprzeczającym możliwościom maszyn matematycznych, tak jak i nie zaprzecza ono podobnym możliwościom człowieka. Człowiek w tym tylko jest „lepszy" od maszyny Turinga, że sformułował twierdzenie o niezupełności, poza tym ich inteligencja (jak zakłada się w teoriach sztucznej inteligencji) jest porównywalna i w zasadzie daleko wykracza poza dowodzenie prawdziwości pojedynczego twierdzenia w ramach zamkniętego niesprzecznego systemu.

Maszyny Turinga i ludzie są zatem zrównani wobec swych możliwości poznawczych ze względu na twierdzenie Gödla; to zrównanie jest jednak w istocie ich ograniczeniem, konkludują niektórzy teoretycy. Gdy oba systemy (układy) poznawcze potrafił uporać się częściowo z twierdzeniem Gödla, stosując inne niż obliczalne (algorytmiczne) metody dowodzenia, to w czynności tej nie różnią się jednak jakościowo. Opinię tą wyraża M. Apter, pisząc:

> Z pewnością jest prawdą, że zarówno ludzie, jak i maszyny są przedmiotem twierdzenia Gödla w tym zakresie, w jakim funkcjonują jako układy formalne. (...) Zarówno ludzie, jak i maszyny mogą w pewnych warunkach przezwyciężyć ograniczenia, o jakich mówi twierdzenie Gödla, tolerując zdarzające się niekonsekwencje i błędy, które są prawie nieuniknione przy zastosowaniu metod heurystycznych, a w gruncie rzeczy jedne i drugie podlegają ograniczeniom

> narzucanym przez to twierdzenie. [M. Apter, *Komputery a psychika. Symulacja zachowania*, s. 115].

Nawet stosowanie metod heurystycznych nie wyróżnia (uprzywilejowuje) człowieka wobec maszyny matematycznej, która tą metodę tylko symuluje; wszelkie podobieństwa są w ostateczności dowodem ograniczeń umysłu ludzkiego.

Nie wszyscy badacze maszyn matematycznych wyrażają powyższy pogląd co do zasadniczego podobieństwa maszyn i umysłów, nie wszyscy żywią pesymizm co do ograniczonych możliwości poznawczych człowieka. E. Nagel i J.R. Newman mówią wprawdzie, że twierdzenie Gödla wskazuje na pewną ograniczoność maszyn matematycznych w ogóle, komputerów w szczególności, w dowodzeniu prawdziwości twierdzeń systemu aksjomatycznego, lecz nie wyciągają pesymistycznych wniosków co do ograniczonych możliwości umysłu ludzkiego. Twierdzenia Gödla należy interpretować w tej materii ani pesymistycznie, ani mistycznie. Odkrycie, że istnieją prawdy, dla których nie ma dowodu posiadającego reprezentację w ramach arytmetyki nie oznacza, że nie można w ogóle skonstruować ściśle finitystycznego dowodu prawdziwości danego twierdzenia. Jest to w zasięgu możliwości człowieka, nie ma tu żadnych „nieprzekraczalnych granic ludzkiego rozumu", takiego wniosku twierdzenie Gödla nie implikuje.

> Dowodzi natomiast – piszą – że działalność intelektu nie została dotąd i nie może zostać nigdy w pełni sformalizowana, że nowe zasady dowodzenia czekać będą na odkrycie. (...) Twierdzenie to wskazuje natomiast, że struktura i działalność umysłu ludzkiego jest daleko bardziej złożona i subtelna niż budowa i sposób funkcjonowania którejkolwiek z maszyn, jakie dziś potrafimy zaprojektować. Dzieło Gödla jest znakomitym przykładem tej złożoności i subtelności. Skłania ono nie do zwątpienia, lecz do wzmożonej ufności w potęgę twórczego umysłu. [1966, s. 71].

Nagel i Newman zakładają, że w ramach finitystycznej interpretacji matematyki dowód taki jest w zakresie możliwości człowieka, jednak nie musi (ale i nie może) być maszynowo wykonany. Kwestia przeprowadzenia takich dowodów jest zatem wciąż otwarta.

W sprawie porównania możliwości poznawczych (obliczeń) maszyny matematycznej i człowieka wypowiedział się również sam Turing. W istocie jego zdanie w tej sprawie wywołało wielką dyskusję w ramach różnych teorii sztucznej inteligencji, ukierunkowało jednak uwagę wielu teoretyków nadmiernie w jedną stronę. Wprawdzie przyznał on, że pytanie „czy maszyny mogą myśleć?" jest nazbyt nieokreślone, to jednak wielokrotnie (zwłaszcza w wypowiedziach i tekstach po wojnie) dał podstawy do takiego właśnie ogólnego, niepoprawnego stawiania problemu; sformułował również kilka trafnych uwag na temat natury ludzkiego umysłu i jego matematycznego modelu.

Najpełniejszym wyrazem stanowiska Turinga w powyższej kwestii jest jego artykuł z 1950 roku pt. *Maszyna licząca a inteligencja*, zawierający argument w postaci tzw. gry w udawanie. Pierwsza część artykułu najbardziej przyczyniła się (nie do końca zresztą w zgodzie z intencją autora) do rozpowszechnienia się przekonania, że cyfrowe komputery mogą być nierozróżnialne w stosunku do pewnych działań (udzielania odpowiedzi na pytania) człowieka. Turing przyznał, że wobec faktu, że każda konkretna maszyna matematyczna o stanach nieciągłych nie może wykonać pewnych działań (co zostało potwierdzone tzw. tezą Turinga-Churcha) wnioskowanie, że umysł ludzki nie podlega takim ograniczeniom nie zostało poparte żadnym dowodem. Nie daje to zresztą żadnej przewagi człowiekowi wobec maszyny, bowiem niemożność jednej maszyny może być pokonana przez maszynę drugą.

> Ostatecznie – pisze Turing – ową wyższość możemy odczuwać w stosunku do tej konkretnej maszyny, nad którą odnosimy nasze skromne zwycięstwo. Zwycięstwo takie nad wszystkimi maszynami jednocześnie w ogóle nie wchodzi w grę. Krótko więc mówiąc, jeśli nawet człowiek okazuje się bystrzejszy od jakiejkolwiek istniejącej maszyny, to powstać mogą inne, jeszcze bystrzejsze maszyny itd. [A. Turing, *Maszyny liczące a inteligencja*, s. 284].

Możliwość porównania cyfrowej maszyny liczącej o stanach dyskretnych z działaniem umysłu (także mózgu) zasadniczo ciągłego Turing proponował rozważyć na poziomie nie reguł działania (zachowania, reguł określających pracę sprzętu), lecz na poziomie reguł wnioskowania maszyny, jej programu. Maszyną w pełni symulującą pracę umysłu, w tym głównie jego pozaformalne operacje, mogłaby być maszyna z elementami losowymi, z pewnym wbudowanym w jej program odstępstwem od reguł.

> Zachowanie inteligentne wiąże się pewnie z jakimś odstępstwem od zachowania całkowicie zdyscyplinowanego, właściwego przy przeprowadzaniu obliczeń, ale na tyle niewielkim, żeby nie było to źródłem działania na chybił trafił bądź jałowych zapętleń. (...) Należy sądzić, że maszyna ucząca się powinna należeć do maszyn z elementem losowym. Działanie losowe jest dobrą metodą poszukiwania rozwiązań pewnych problemów. [A. Turing, *Maszyny liczące a inteligencja*, s. 298].

W raporcie opisującym ACE, będącym w pełni prototypem komputera, Turing zawarł (w odpowiedzi na pytanie „jak daleko jest w zasadzie możliwe, aby maszyna licząca symulowała ludzkie czynności?") następującą jeszcze uwagę:

> Istnieje wiele twierdzeń zakładających prawie dokładnie, że jeśli od maszyny oczekuje się nieomylności, to nie może ona być jednocześnie inteligentna. Lecz twierdzenia te nie mówią niczego o tym, jak bardzo może ujawnić się inteligencja, jeśli tylko maszyna nie posiada pretensji do nieomylności. [M. Davies, *The Undecidable: Basic Papers on Undecidable Propositions*, s. 170].

W ostateczności – konkludował Turing niejako wbrew założeniu o niemożności podania rozstrzygającego wyniku (kto jest kto) w grze w naśladownictwo między człowiekiem a komputerem – należy oczekiwać, że maszyny cyfrowe będą raczej rywalizowały z człowiekiem w pewnych czynnościach intelektualnych niż będą całkowicie jego w tym naśladowały lub zastępowały.

Powyższe uwagi Turinga i innych autorów wskazują zgodnie na znaczenie twierdzenia Gödla (jego epistemologicznego znaczenia) dla analizy umysłu. Zakłada się w nich (najczęściej *implicite*), że twierdzenie to ma także charakter empiryczny i nie stosuje się wyłącznie do formalnych systemów wiedzy; różnica pojawia się dopiero w opiniach na temat przewagi czy niedostatku umysłu wobec maszyny Turinga. Uwagi te formułowane są jednakże w ramach jednego fundamentalnego (nie w pełni uświadamianego ze względu na poważne ograniczenia) założenia, że **umysł jest względnie wyizolowanym aspektem działań człowieka** oraz, że **istotą umysłu są czynności intelektualne.** Umysł traktowany jest tu bardzo ogólnie, bez zróżnicowania na rodzaje czynności poznawczych jakich jego funkcjonowanie wymaga. Zakłada się, że maszyna Turinga symuluje działanie abstrakcyjnego umysłu jako takiego i zasadniczo w całości. To zaś jest wysoce dyskusyjne.

5. Jaki powinien być maszynowy model umysłu? Zarys problemu

Kwestię sformułowaną w tytule niniejszego paragrafu można wyrazić w innych jeszcze pytaniach. Jaka maszyna (maszyny) może być modelem umysłu ludzkiego, co można wyrazić również inaczej – jaka maszyna (maszyny) może symulować jego działanie? Z pytaniem tym łączy się jeszcze jedno, ogólniejsze – czy możliwa jest ogólna (jedna) teoria umysłu (inteligencji), która by opisywała i wyjaśniała szerokie spektrum czynności poznawczych i praktycznych człowieka i pewnej określonej grupy maszyn? Wydaje się, że wprawdzie pytania te można postawić i rozstrzygać niezależnie od siebie, to jednak w odpowiedzi na pytanie pierwsze należałoby uwzględnić pewne ustalenia wynikające z prób odpowiedzi na pytanie drugie.

Ogólna teoria umysłu T U musiałaby spełnić następujące warunki: (1) dla bardzo szerokiej klasy **podmiotów** P jak człowiek, maszyna matematyczna, czy każdy układ cybernetyczny należałoby wyznaczyć (2) względnie szeroką **dziedzinę poznawczą** D z wyróżnioną podklasą (3) **czynności dowodzenia** prawdziwości twierdzeń d/p (lub inaczej mówiąc, używania predykatu „prawdziwy” w odniesieniu do takich wyrażeń językowych jak zdania, wypowiedzi itp.); oraz (4) uwzględnić epistemologiczny sens twierdzenia Gödla G (prawdziwość nie jest tożsama z dowodem, istnieją procedury niealgorytmiczne). Formuła T U (P, D, d/p, G) znaczyłaby wówczas „teorię umysłu dla takich podmiotów jak ludzie czy maszyny matematyczne w ich (ograniczonych przez twierdzenie Gödla) czynnościach dowodzenia prawdziwości twierdzeń”. Perspektywy na zbudowanie takiej teorii, o której marzy wielu teoretyków sztucznej inteligencji, są raczej ograniczone. Krytycznie o takiej możliwości (w odniesieniu do czynności uczenia się języka, nabywania kompetencji językowych) wypowiedzieli się zgodnie J. Piaget, N. Chomsky i H. Putnam (por. Rosner, 1995, ss. 256-260), argumentując, że perspektywy jej zbudowania są równie mało prawdopodobne jak próby (dotychczas nie udane) uzyskania „ogólnej teorii wzrostu”. Niemniej teoria taka jest szczególnym wyzwaniem intelektualnym i wydaje się, że już częściowe zrealizowanie któregoś z jej punktów można byłoby uznać za spory sukces.

Złożoność powyższego zadania polega na zdefiniowaniu „dziedziny poznawczej”, w obrębie której spełniane mają być czynności dowodzenia prawdziwości; jest to najbardziej trudny do określenia z warunków ogólnej teorii (modelu) umysłu. Najczęściej zakłada się, że dziedziną tą ma być matematyka, ściślej, aksjomatyczne systemy finitystycznego dowodzenia prawdziwości jej wyrażeń. Ale już twierdzenie Gödla i teza Turinga-Churcha pokazują, że podklasa czynności dowodzenia prawdziwości nie sprowadza się do jednej tylko procedury, lecz rozpada się na dwie jeszcze podklasy: dowodzenie algorytmiczne i niealgorytmiczne. Mając to na uwadze należałoby zatem uzupełnić treść ogólnej teorii umysłu również o nieformalne, infinitystyczne, niezupełne (niespójne) obszary wiedzy i poznania, także o czynności niealgorytmicznego dowodzenia prawdziwości, szerzej, wartościowania wiedzy wyrażonej nie tylko w postaci propozycjonalnej (twierdzeń, zdań), lecz również aktów mowy, sadów.

Konieczność poszerzenia dziedziny poznania poza formalne systemy aksjomatyczne i uwzględnienia pozapropozycjonalnych jednostek wiedzy powoduje, że ogólna teoria (model) umysłu musiałaby uwzględnić rzeczywiste sytuacje niealgorytmicznego, heurystycznego, twórczego rozwiązywania (zarówno przez człowieka, jak i maszyny) szerokiej klasy problemów poznawczych. W istocie trzeba uwzględnić poza formalnymi procedurami dowodzenia także rzeczywiste czynności pozaformalnego postępowania wobec różnych problemów poznawczych. Dziedziną tych czynności jest **heurystyka**, której cele i warunki zostały określone przez G. Polya:

> Podstawą na której buduje się heurystykę, musi być doświadczenie w rozwiązywaniu zadań i doświadczenie w obserwowaniu innych ludzi rozwiązujących zadania. Nie można przy tym lekceważyć żadnego rodzaju zadań. Należy wyszukiwać wspólne cechy sposobów traktowania wszystkich rodzajów zadań. [1964, ss. 135-136].

Przy budowie heurystyki należy uwzględnić tak logiczny, jak i psychologiczny, pedagogiczny jej aspekt. Tą dziedzinę H charakteryzuje zasadniczo wyróżniona klasa czynności **dokonywania odkryć** d/o, która może być ujęta w szereg reguł, lecz nie jest ściśle sformalizowana w postaci skończonych procedur. Heurystyka jest uzupełnieniem metod algorytmicznych, współdziała z nimi; biorąc pod uwagę status teorii algorytmów i heurystyk można by powiedzieć, że pierwsza jest aprioryczna, druga aposterioryczna. Współczesne rozumienie heurystyki wykracza poza znaczenie nadane przez Polya (jego metody dotyczyły głównie odkrywania i wymyślania rozwiązań w ogóle, dopuszczały również „działanie po omacku") i odnosi się do specyfiki konkretnej dziedziny, w której dany problem się pojawia, polega głównie na poprawianiu metod i strategii już istniejących[5].

Uwzględniając powyższe, teorię (model) umysłu należałoby wyrazić obecnie w formule poszerzonej: T U (P, H, d/p & d/o, G), gdzie algorytmiczne procedury dowodzenia prawd byłyby tylko szczególnym przypadkiem klasy szerszej – heurezy, czyli dokonywania odkryć; także wobec nich obowiązywałby sens twierdzenia Gödlowskiego (w znaczeniu, w jakim mówił Nagel i Newman). Teoria taka nie mogłaby jednak zawierać „niezawodnych reguł (algorytmów) **wszystkich przyszłych** problemów"[6]. Teoria o niealgorytmiczności radzenia sobie z określonymi sytuacjami poznawczymi nie może bowiem sama być sumą algorytmów dlatego, że niemożliwością poznawczą jest przewidzenie (co najmniej częściowe zalgorytmizowanie, obliczenie) wszystkich problemów i czynności poznawczych, nawet jeśli znane są (częściowe) reguły radzenia sobie z (względnie) szeroką klasą problemów.

Model powyższy został praktycznie (w wąskim zakresie) zrealizowany w postaci programu komputerowego pod nazwą Maszyny do Teorii Logiki (Logic Theory Machine), napisanego w 1956 roku przez A. Newella, J. C. Shawa i H. Simona[7]. Jest to pierwszy heurystyczny program całkowicie zrealizowany na maszynie cyfrowej, który w zamyśle autorów miał służyć do symulowania czynności rozwiązywania bardzo szerokiej klasy problemów, jak dowodzenie twierdzeń matematycznych (niektórych z *Prinicpia Mathematica* Russella i Whiteheada), gry w szachy, a także rozumienia języka potocznego. Kolejne wersje (np. General Problem Solver) powyższej maszyny miały tą samą strategię działania: maszyna przekształca wejściowe wyrażenia (aksjomaty, wyrażenia już udowodnione) generując w oparciu o rachunek algebraiczny ciągi dowodowe. Zbiory wygenerowanych ciągów dowodowych mogą jednak wzrastać według zasad eksplozji kombinatorycznej; dla niektórych twierdzeń wyjściowych znalezienie dowodu (odpowiedzi na pytanie, rozwiązanie danego problemu) może być przez to niewykonalne, tj. zająć zbyt dużo czasu czy wymagać zbyt dużego kosztu obliczeń.

W podstawowej części swojego działania maszyny logiczne Newella, Shawa i Simona mają zatem te same ograniczenia, na jakie napotyka maszyna Turinga; są również pewne nowe

---

[5] Por. Bolc, Cytowski, *Metody przeszukiwania heurystycznego*, ss. 9-10.
[6] Por. Z. Cackowski, *Człowiek jako podmiot działania praktycznego i poznawczego*, s. 439.
[7] Por. *Maszyny matematyczne i myślenie*, red. E. A. Feigenbaum, J. Feldman, 1972, ss. 118-144.

rozwiązania. Część algorytmiczna współdziała bowiem z częścią heurystyczną, której rola sprowadza się do „inteligentnej" oceny, w istocie redukcji generowanych ciągów dowodowych. Generowane algorytmicznie stanowią one klasę podproblemów dla problemu głównego rozważanego heurystycznie. Ta algorytmiczno-heurystyczna procedura wymaga zatem zasadniczo nowej strategii działania.

> Przeprowadzania złożonych procesów decyzyjnych w środowisku potencjalnie nieskończonym i wymykającym się spod kontroli. [Gelernter w: *Maszyny matematyczne i myślenie*, s. 145].

Radzi sobie z tym tzw. filtr heurystyczny, który selekcjonuje i wybiera właściwe ciągi dowodowe, określa i szacuje koszty obliczania, wyznacza prawdopodobny kierunek rozwiązania głównego problemu. Ta strategia ma swoje zalety i wady: gwarantuje skuteczne rozwiązanie problemu kosztem rezygnacji z optymalności końcowego wyniku. Optymalność oznacza zazwyczaj wybór najlepszego (efektywnego, obliczanego wg określeń Turinga) ciągu dowodowego, lecz z racji wzrostu czasu i kosztów obliczeń jest to niekiedy nieopłacalne. Oparta na prawdopodobieństwie heureza, łącząca się z ryzykiem poznawczym, jest nie tyle alternatywą dla algorytmicznej procedury, co jej dopełnieniem.

Maszynę algorytmiczno-heurystyczną można potraktować jako model powstały przez rozwiniecie i uzupełnienie zasadniczych założeń maszyny Turinga, ale także (co najważniejsze) jako utworzony w oparciu o obserwację i uogólnienie faktycznych procedur (eksperymentalnie przeprowadzanych w laboratoryjnych warunkach) rozwiązywania konkretnych zadań poznawczych. Z tego względu model ten można uznać za lepsze przybliżenie (symulowanie) szerszej grupy czynności poznawczych i praktycznych człowieka. W większym stopniu uwzględnia on konkretność (różnorodność, odmienność) modelowanych przypadków, w mniejszym zaś ogólność (abstrakcyjność, uniwersalność) inteligencji człowieka. Wyrazem tego są prace prowadzone w ramach badań nad tzw. sieciami neuronowymi, algorytmami genetycznymi i ewolucyjnymi, także systemami eksperckimi – nową generacją programów, które w zamyśle twórców są dalszym i lepszym modelem umysłu ludzkiego.

Czy modele (teorie) te są naprawdę poprawnymi prezentacjami umysłu ludzkiego? Wydaje się, że odpowiedź jest wciąż ta sama – nie. Gdy klasyczna maszyna Turinga symuluje zaledwie wąską klasę czynności dowodzenia jakie człowiek (w istocie matematyk, i to nie każdy) przeprowadza wobec systemów formalnych, to i tak poza jej modelem pozostają operacje pozaformalnego „wglądu w prawdę", o których R. Penrose pisze następująco:

> Procedury umysłowe, które służą matematykom do rozstrzygania, czy dane zdanie jest fałszywe, czy prawdziwe, nie wynikają z procedur pewnego systemu formalnego. (...) Prawda matematyczna wykracza poza ludzkie konstrukcje. [R. Penrose, *Nowy umysł cesarza: o komputerach, umyśle i prawach fizyki*, ss. 132-134].

Jego argumenty na rzecz intuicyjnego, Platońskiego wglądu w absolutny świat matematyki, bez mała kontemplacyjne odkrywanie prawd, są kontrowersyjne; nie są zresztą jedyną interpretacją dokonywania odkryć naukowych w matematyce. Uwzględnienie w modelu maszyny algorytmiczno-heurystycznej szerszego spektrum czynności poznawczych, w tym dodatkowo probabilistycznych, losowych procedur dokonywania odkryć też nie wydaje się innym jakościowo rozwiązaniem. Prace teoretyczne i konstruktorskie w dziedzinie sztucznej inteligencji są próbami wymodelowania maszynowego działania umysłu ludzkiego i symulowania go na maszynach liczących. Są realizacją pomysłu i oczekiwań samego Turinga, stopień ich zaawansowania i uzyskiwane efekty są jednak przedmiotem rozbieżnych opinii.

Modelowane są zasadniczo tylko **pojedyncze czynności poznawcze** jak rozpoznawanie (monozmysłowe) ściśle wyróżnionych ze środowiska cech obiektów, obrazów, dźwięków czy mowy. Poza możliwościami symulacji pozostaje wciąż kompleks praktycznych czynności poznawczych człowieka (o wiele lepiej symulowana jest receptoryka niż motoryka), których umysł jest funkcją w stopniu nie mniejszym niż zmysłów. Nawet samouczące się sieci neuronowe (poprawnie mówiąc, neuropodobne), nad którymi przeprowadza się niezwykle rozwinięte i intensywne badania, nie są satysfakcjonującym modelem umysłu człowieka, gdyż są raczej bardzo przybliżonym modelem nawet nie mózgu całego, lecz działania jego elementarnych modułów – neuronów i ich lokalnych synaptycznych połączeń.

> Sieć neuronowa jest bardzo uproszczonym modelem mózgu. Składa się ona z dużej liczby (od kilkuset do kilkudziesięciu tysięcy) elementów przetwarzających informację. Elementy te nazywane są neuronami, chociaż w stosunku do rzeczywistych komórek neuronowych ich funkcje są bardzo uproszczone, by nie powiedzieć – sprymitywizowane. [Tadeusiewicz 1995, ss. 18-19].

Sieć taka modeluje zatem nie umysł i jego czynności, lecz fragmentarycznie zbadane (wciąż niewystarczająco) procesy mózgowe, które im towarzyszą, które je warunkują. Dlatego też nie jest to jeszcze poprawny (bogaty, adekwatny) model człowieka, mimo że niektóre zasady działania sieci neuropodobnej (implementowanej na sprzęcie komputerowym) określone zostają mianem (raczej metaforą) „uczenia się” przez analogię do niektórych czynności człowieka.

***

Podsumowując powyższe maszynowe modele (teorie) człowieka trzeba podkreślić, że ich wspólną i charakterystyczną cechą jest **atomizujące, selektywne i jednostronne ujmowanie czynności poznawczych.** W poszczególnych przypadkach symulowane są przez maszyny (komputery cyfrowe) takie jednostkowe działania jak: operacje dowodzenia, stosowanie reguł danej gry, przekład między językami, rozpoznawanie obiektów, heurystyczne podejmowanie decyzji, modyfikacja (uczenie się) nabytych umiejętności itp. To spektrum – wciąż poszerzane i doskonalone – teoretycznych modeli (programów) i skutecznych implementacji na maszynach (robotach) uznaje się za adekwatny obraz ludzkiego poznania i umysłu. Zakładając nawet, że ilość, precyzja i efektywność programów symulujących poznanie i umysł będzie wzrastać, to i tak nie będą one adekwatnymi modelami (teoriami), gdy poza ich zakresem, ale również możliwościami, pozostanie to, co stanowi istotę ludzkiej aktywności poznawczej – **realizowanie się wobec konkretnego środowiska, w oparciu o przedmioty (narzędzia, znaki, symbole), ze względu na środki i cele.** Aktywności tej nie charakteryzuje w całości żadna jedna, szczególna reguła. Nie jest ona ani zupełnie zalgorytmizowana, ani całkowicie chaotyczna, przypadkowa czy losowa; jej istota wyczerpuje się w spektrum przypadków od skrajnego nieuporządkowania, chaosu po próby jego uporządkowania, zalgorytmizowania, zawsze częściowego[8]. Nie sprowadza się ona ponadto do jednorazowych aktów układających się w ciągi dyskretnych, skokowo przebiegających elementów. Tylko stosunkowo nieliczne działania poznawcze i praktyczne człowieka można opisać (modelować i symulować maszynowo) w kategoriach funkcji rekurencyjnych, obliczalnych. Swoistą

[8] Z. Cackowski, *Rozum między chaosem a „Dniem Siódmym” porządku*, UMCS, Lublin 1997, ss. 65-109.

„regułą" działania człowieka jest raczej to, że nie podlega ono ani wyłącznie, ani też najczęściej regułom dającym się ściśle opisać i obliczyć.

Działanie człowieka, w przeciwieństwie do działania większości maszyn (w tym cyfrowych komputerów), charakteryzuje się posiadaniem (ale też nie w każdym przypadku) takich reguł, które są **immanentnie zawarte w działaniu**; są one w jego trakcie zmieniane, wtedy też dopiero są tworzone. Z kolei pewna część reguł istnieje przed działaniem, jest powiązana ze sobą, układa się w program działania. Można zatem rozróżnić w działaniu człowieka dwa rodzaje reguł: **regulatywne**, które określają istniejące uprzednio i niezależne od nich działanie człowieka (np. reguły zachowania się przy stole), które są wtórne i przypadkowe; oraz **konstytutywne**, które ustanawiają dopiero jakiś rodzaj działania, powołują go do życia (np. reguły gry w szachy), które są umowne i przez to jednoznaczne[9].

Analizując (modelując) działanie człowieka należy uwzględnić jeszcze reguły (prawidłowości) jakim podlega jego ciało, w tym układ nerwowy i procesy mózgowe.

> W istocie rzeczy powinno by się mówić: **procesy** mózgowe i **działania** umysłowe. [Cackowski 1997, s. 95].

Procesy nerwowe są czynnikiem determinującym behawioralne reakcje człowieka ale nie są czynnikiem jedynym, także nie głównym. Działanie człowieka jest bowiem zasadniczo warunkowane zewnętrznymi rzeczami, obiektami środowiska, ich fizycznym oddziaływaniem na organizm. Ponadto w charakterze czynnika warunkującego występują idealne (niezmysłowe, pojęciowe) treści doświadczenia, motywy, cele i intencje. Dopiero konglomerat tych czynników – immanentnych i zewnętrznych reguł, procesów cielesnych i intencji – stanowi o całości działania. Modelowanie i symulowanie któregokolwiek z tych aspektów i całości działania człowieka winno tę złożoność uwzględniać. Ale czy istnieje maszyna będąca modelem takiej całości, czy możliwa byłaby na niej jej symulacja?

***

Do wymodelowania, zaprogramowania, symulowania na jakiejś maszynie (maszyna Turinga musiałaby być jej częścią) pozostaje zatem nie tylko cielesne (procesualne, fizjologiczne) uwarunkowanie działania, lecz zasadniczo **środowisko działania człowieka i jego współdziałanie z innymi ludźmi**. Jest to zagadnienie o kapitalnym znaczeniu, gdyż jakakolwiek czynność praktyczno-poznawcza jednostki ma swoje uwarunkowanie – także znaczenie – w faktycznych relacjach i uwikłaniach z przedmiotami środowiska i innymi ludźmi. Ta oczywista prawda oznacza jednak w odniesieniu do tytułowego zagadnienia istotną komplikacją i trudność.

Czy istnieje taka maszyna, która byłaby modelem (teorią) człowieka działającego wobec rzeczy i współpracującego z innymi ludźmi, a nie tylko wykonującego proste operacje dowodzenia, rozpoznawania obiektów, przekładu jednego języka na drugi, rekonstrukcji dokonanych już odkryć itp.? Jest oczywiste, że jakakolwiek konkretna maszyna Turinga nie jest takim modelem, nie jest nim żaden z dotychczasowych komputerów cyfrowych. Turing zakładać jednakże istnienie uniwersalnej maszyny, która może sumować moc obliczeniową

[9] Por. Searle, *Umysł, mózg i nauka*, PWN, 1995, ss. 52-63; J. Bobryk, *Akty świadomości i procesy poznawcze*, ss. 106-113.

każdej maszyny konkretnej i symulować jej działanie. I chociaż sumowanie obliczeń to tylko zmiana ilościowa możliwości, to może jednak możliwa jest do wyobrażenia jako uniwersalna maszyna do modelowania bogactwa człowieka?

Aby rozważyć możliwość istnienia prawdziwie uniwersalnej (w szerszym znaczeniu) maszyny (modelu) człowieka należałoby dokonać paru ważnych modyfikacji w budowie i zasadach działania maszyny Turinga. Ich skrótowy, wstępny (do rozwinięcia) rejestr wyglądałby następująco:

Po pierwsze, zbiór stanów maszyny (zawartych w jakiejś jednostce centralnej, układzie sterowniczym) określający ilość i rodzaj wykonywanych operacji musi być w zasadzie **nieograniczony,** bliski nieskończoności. Maszyna musi byś wystarczająco bogata w budowie, zróżnicowana i rozbudowana, zawierają liczne części współpracujące między sobą, sterowane przez układ kierowniczy. Taka maszyna musi być w stanie wykonać względnie dużo zadań, które może napotkać. W rejestrze możliwości maszyny muszą znajdować się (w skończonej liczbie) stany stałe, w jakich maszyna może działać efektywnie oraz stany, które tylko potencjalnie zawarte są w jej budowie (konstrukcji), które mogą się zaktywizować dopiero w danym momencie; w tym drugim przypadku można byłoby mówić o nieskończoności maszyny. Maszyna musi wykonać potencjalnie o wiele więcej czynności niż może wykonać w którymkolwiek z zarejestrowanych (skonstruowanych) stanów, więcej niż wykonuje w danym trybie pracy.

Po drugie, program jej działania musi się łatwo nie tylko wymieniać, ale również rozbudowywać. Nadto musi istnieć możliwość jego **zmiany w trakcie wykonywania** (to najdalej idąca modyfikacja w stosunku do założeń Turinga). Maszyna musi uczyć się poprzez kolejne modyfikacje wykonywanego programu. Program jako zbiór reguł musi być nie tylko początkiem działania maszyny (wyznaczać jego kierunek), lecz także – i przede wszystkim – treścią tego działania (być wyznaczony przez nie); program musi nie tylko konstytuować (determinować) działanie maszyny, ale także być przez nie regulowany.

Po trzecie, taśma (która oznacza nieskończone możliwości operowania przez maszynę znakami, dowolnymi danymi, każdą informacją) musi być w istotny sposób **wewnętrznie zdeterminowana**. Turing zakładał, że choć maszyna operuje wobec nieskończonej taśmy, to jednak ma do czynienia ze skończonym zbiorem znaków przyjętych konwencjonalnie. Wobec konkretnego znaku (0, 1 lub braku znaku) maszyna wykonuje operacje zasadniczo zdeterminowane którymś ze stanów, w jakim się znajduje (zapisanym w rejestrze); to co maszyna „wie" jest bardziej określone przez jej stan wewnętrzny niż zewnętrzny (charakterystykę taśmy). Ograniczenie to musi być zmodyfikowane w kierunku zasadniczej determinacji samej taśmy, tj. współwyznaczania "wiedzy" maszyny przez informację z taśmy na równi z jednostką centralną maszyny. To co jest zapisem na taśmie (każda informacja dowolnie zakodowana) powinno być nie tylko operacyjnie (biernie) obliczone na bieżąco i zapisane w pamięci maszyny, ale także musi determinować stany maszyny, by zwrotnie wpływać na kolejną przyjmowaną informację. Słowem, informacja miniona i bieżąca muszą współdziałać w oparciu o mechanizm sprzężenia zwrotnego negatywnego.

Po czwarte, powyższe pociąga za sobą konieczność modyfikacji zasadniczej – odstępstwo od reguły przerywistego (dyskretnego) działania na rzecz **ciągłości** stanów maszyny. Aby w pełni, w całości jakakolwiek maszyna modelowała myślenie człowieka musi ona być także maszyną stanów ciągłych (a nie tylko maszyną stanów dyskretnych). Ciągłość, nieprzerywistość jest bowiem konstytutywną cechą myślenia człowieka, którego model wymaga odejścia od ściśle deterministycznej organizacji. W istocie modelować trzeba ciągło-przerywisty charakter myślenia ludzkiego, które na wielu poziomach, w zależności od użytych środków przybiera którąś z tych własności. Dla zrealizowania zasady ciągłości (co najmniej jej imitacji) maszyna taka powinna ponadto wykonywać więcej niż jedną operację na raz, przez co musiałaby znajdować się w różnych stanach jednocześnie i operować większą ilością informacji. Jej działanie musiałoby być równolegle i wielokierunkowe. Z tego względu pełny model umysłu ludzkiego musiałby być maszyną tak samo cyfrową, jak i analogową, działającą tak szeregowo, jak i równolegle.

Literatura:

[1] M. Apter, *Komputery a psychika. Symulacja zachowania*, PWN, Warszawa 1973.

[2] J. Bobryk, *Akty świadomości i procesy poznawcze*, Wyd. Leopoldinum, Wrocław 1996.

[3] L. Bolc, J. Cytowski, *Metody przeszukiwania heurystycznego*, PWN, Warszawa 1989, t. 1 i 2.

[4] Z. Cackowski, *Człowiek jako podmiot działania praktycznego i poznawczego*, KiW, Warszawa 1979.

[5] Z. Cackowski, *Rozum między chaosem a „Dniem Siódmym" porządku*, UMCS, Lublin 1997.

[6] M. Davies, The Undecidable: Basic Papers on Undecidable Propositions, w: *Unsolvable Problems and Computable Functions*, red. M. Davies, Raven Press, New York 1965.

[7] M. Davies, Mathematical Logic and the Origin of Modern Computers, w: R. Herken, *The Universal Turing Machine...*, dz. cyt. ss. 149-174.

[8] R. Gandy, The Confluence of Ideas in 1936, w: R. Herken, *The Universal Turing Machine...*, dz. cyt. ss. 55-111.

[9] *Maszyny matematyczne i myślenie*, red. E.A. Feigenbaum, J. Feldman, PWN, Warszawa 1972.

[10] R. Penrose, *Makroświat, mikroświat i ludzki umysł*, Prószyński i S-ka, 1997.

[11] A. Turing, *Maszyny liczące a inteligencja*, w: *Filozofia umysłu*, red. B. Chwedeńczuk, Warszawa 1995.