

Tytuł: **Funkcjonalizm jako filozoficzna podstawa teorii Sztucznej Inteligencji**

Autor: Piotr Kołodziejczyk / pkolodziejczyk@interia.pl

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

Data publikacji: 04 VII 2002

1. Uwagi wstępne

Jeśli badania nad sztuczną inteligencją pojmuje się nie tylko jako zbiór praktycznych działań mających na celu konstrukcję systemów nieodróżnialnych (lub porównywalnych) od człowieka pod względem możliwości dokonywania operacji poznawczych, ale również jako pewną teorię umysłu, to wydaje się, iż teoria ta winna opierać się na pewnych podstawach filozoficznych¹. Podstawy te mogą bowiem zagwarantować oblicze koherencji systemowej teorii Sztucznej Inteligencji. Synteza wyników uzyskanych przez dyscypliny zajmujące się AI, refleksja nad metodologiczną prawomocnością uzyskanych w ich ramach wyników oraz próba sprecyzowania zakresu stosowalności podstawowych dla teorii Sztucznej Inteligencji pojęć, sprawia, iż filozoficzna refleksja nad statusem AI jawić się może jako teoretycznie doniosła:

Należy podkreślić – Pisze Zdzisław Chlewiński. – że w wielu pytaniach stawianych przez twórców modeli umysłu (i badaczy Sztucznej Inteligencji, przyp. P. K.) kryje się, pomimo pozorów nowości, problematyka obecna w dyskusjach filozofów, a także przedstawicieli innych nauk. Chodzi tu m.in. o takie zagadnienia jak: mózg-umysł, świadomość-nieświadomość, redukcjonizm, natura intencjonalności, czy człowiek jest maszyną Turinga? itp..²

W ramach prac nad sztuczną inteligencją, jako „oficjalną” filozofię przyjmuje się stanowisko funkcjonalistyczne. Wynika to, jak sądzę, ze zgodności głównych tez tego stanowiska z twierdzeniami sformułowanymi w ramach innych dyscyplin oscylujących wokół tematyki Sztucznej Inteligencji. Podobnie jak owe dyscypliny (np. psychologia behawiorystyczna), funkcjonalizm proponuje zastosowanie strategii redukcjonistycznej w badaniu stanów (zjawisk) mentalnych. Strategia funkcjonalistyczna jest zaś konsekwencją asymilacji wniosków płynących z obliczeniowej (komputacyjnej) teorii umysłu. Mając tą konstatację na uwadze, w poniższych rozważaniach podejmuję próbę analizy głównych założeń obliczeniowej teorii umysłu i jej wpływu na ukształtowanie się filozofii funkcjonalistycznej.

¹ Por. J. Fodor, *Jak grać w reprezentacje umysłowe*, ss. 17-49, w: *Modele umysłu*, red. Z. Chlewiński, PWN, Warszawa 1999, s. 25.

² Z. Chlewiński, *Wprowadzenie*, s. 14, w: tenże, *Modele...*, ss. 7-15.

2. Główne założenia obliczeniowej teorii umysłu

Zdaniem M. Hetmańskiego termin *obliczeniowa teoria umysłu* odnosi się

do wielu koncepcji powstałych i rozwijanych w ramach nurtu kognitywistycznego. Pojawiły się one na pograniczu wspólnych badań nauk, jak logika, informatyka, (...) nauki o poznawaniu, teorie i badania nad sztuczną inteligencją, także lingwistyka i neuronauk.³

Należy jednakże pamiętać o tym, że podstawy komputacyjnej teorii umysłu zostały wyłożone w klasycznej pracy Alana Turinga *Computing Machinery and Intelligence*. W niej to właśnie prekursor badań nad sztuczną inteligencją podjął próbę wykazania, że maszyna matematyczna o stanach nieciągłych (dyskretnych)⁴ posiada identyczne z ludzkimi zdolności w kwestii dokonywania operacji poznawczych (w szczególności – obliczania, wnioskowania i dowodzenia twierdzeń). Turing starał się więc dowieść, iż maszyna ta jest równoważna ludzkiemu umysłowi pod względem zdolności poznawczych⁵.

Założenie o identyczności maszyn cyfrowych i ludzkiego umysłu wynikało z dostrzeżenia analogii pomiędzy procesem modyfikacji instrukcji działania maszyny a ewolucją ludzkiego uczenia się. W jednym z raportów na temat projektu ACE (1947 r.), Turing pisał na ten temat:

Ucząca się maszyna nadal dawałaby wyniki, jakich od niej oczekiwano na początku, lecz dochodziłaby do nich w sposób znacznie bardziej skuteczny. W takim przypadku należałoby przyznać, że rozwój maszyny wykracza poza przewidywania, jakie towarzyszyły wprowadzaniu do niej oryginalnych instrukcji – tak jak uczeń, który wiele nauczył się od swego mistrza, lecz wzbogacił swoją wiedzę także dzięki własnej pracy. Gdy tak się zdarzy, czuję, że nie pozostaje nic innego, jak uznać, że maszyna wykazuje inteligencję.⁶

Postulowanie analogii pomiędzy ludzkim uczeniem się a procedurą programowania maszyn cyfrowych może prowadzić do stwierdzenia, że procesy poznawcze człowieka i maszyny posiadają charakter obliczeniowy. Znaczący to, iż procesy te dają się wyrazić za pomocą skończonego ciągu symboli powiązanych ze sobą za pomocą reguł transformacji zaimplementowanych w strukturze systemu przetwarzającego informacje. Turing przyjął bowiem, iż możliwe jest porównanie

człowieka w trakcie obliczania liczby rzeczywistej do maszyny, która jest w stanie wykonać jedynie skończoną liczbę warunków $q_1, q_2 \dots q_n$, które można określić jako *m-konfiguracje*. Maszyna jest wyposażona w *taśmę* (przypominającą papierową) przechodzącą przez nią i podzieloną na odcinki (zwane *polami*), z których każdy

³ M. Hetmański, *Umysł a maszyny. Krytyka obliczeniowej teorii umysłu*, Wydawnictwo UMCS, Lublin 2000, s. 43. Por. także: P. Coveney, R. Highfield, *Granice złożoności. Poszukiwania porządku w chaotycznym świecie*, tłum. P. Amsterdamski, Wydawnictwo Prószyński i S-ka Warszawa 1997, ss. 169-170.

⁴ Techniczny opis maszyn tego typu można znaleźć np. w artykule B. Bowena i P. Kocury; zob. B. Bowen, P. Kocura, *Implementing conceptual graphs in a RDBMS*, w: *Conceptual Graphs for Knowledge Representation*, red. G. W. Minueau, B. Moulin, J. F. Sowa, Verlag, Berlin-Heidelberg 1993, ss. 106-125

⁵ Stwierdzenie to jest wynikiem eksperymentu myślowego Turinga znanego jako *gra w udawanie [imitation game]*. W grze tej uczestniczy dwóch graczy i sędzie. Jeden z graczy jest człowiekiem, drugi maszyną matematyczną (komputerem). Podczas testu sędzia wysyła na ekrany komputerów, przy których siedzą gracze pewien ciąg pytań. Jeśli maszyna udzieli odpowiedzi nieodróżnialnych od odpowiedzi ludzkich, to jest to, zdaniem Turinga, argument na rzecz twierdzenia, że maszyna myśli.

⁶ A. Hodges, *Enigma. Życie i śmierć Alana Turinga*, tłum. W. Bartol, Wydawnictwo Prószyński i S-ka, Warszawa 2002, s. 299.

zawiera symbol. W dowolnym momencie „w maszynie” obecne jest tylko jedno pole, nazywane *r*-tym i zawierające symbol $S(r)$. Pole to można nazwać *połem odczytywanym*. Symbol w polu odczytywanym można określić jako *symbol odczytywany*. Symbol ten to jedyny symbol, którego maszyna jest „bezpośrednio świadoma”. Maszyna poprzez zmianę swych *m*-konfiguracji może efektywnie zapamiętywać pewne symbole z tych, które odczytywała poprzednio. W danym momencie możliwe zachowanie maszyny jest wyznaczane przez *m*-konfigurację q_n oraz odczytywany symbol $S(r)$.⁷

Co więcej, jeżeli założenie to jest prawdziwe, to prowadzi ono do redukcjonistycznego programu opisu procesów poznawczych danego systemu (człowieka bądź maszyny) w kategoriach logiki formalnej. Bezpośrednia konsekwencja twierdzeń ogłoszonych w *Liczbach obliczalnych* z konieczności implikuje akceptację poglądu, iż zachodzenie danego procesu poznawczego jest warunkowane realizacją określonego algorytmu. Zatem, zarówno działanie ludzkiego umysłu, jak i maszyny cyfrowej polega na operowaniu i przekształcaniu ciągów jednostek syntaktycznych⁸. Stąd zaś Turing wysnuł wniosek, że wszystkie procesy poznawcze są procesami obliczeniowymi. Procedury te charakteryzuje się zaś jako funkcje rekurencyjne⁹.

Wyobraźmy sobie, postuluje Turing, że operacje, których dokonuje maszyna matematyczna zostają rozłożone na szereg *prostych operacji*, które są tak elementarne, iż nie można wyobrazić sobie ich dalszego podziału. Każda z takich operacji polega na pewnej zmianie w obrębie fizycznego systemu składającego się z maszyny i jej taśmy. Potrafimy określić stan tego systemu wtedy i tylko wtedy, gdy znamy kolejność symboli na taśmie, jeśli wiemy, które z nich komputer obserwuje (najprawdopodobniej w określonym porządku) oraz stany umysłu komputera. Możemy przyjąć, że w prostej operacji zmiana ulega nie więcej niż jeden symbol. **Wszystkie inne zmiany mogą być rozłożone na proste zmiany właśnie tego rodzaju** (wytłuszczenie P. K.).¹⁰

Godnym podkreślenia jest fakt, że argumenty Turinga mają zarówno tak matematyczne, jak i psychologiczne podłoże. Analizując powyższą tezę, M. Hetmański wskazuje, iż przyjęty przez autora *Computing Machinery...* wyidealizowany obraz możliwości opisu umysłu za pomocą procedur formalnych jest konsekwencją przyjętego *implicite* programu psychologii behawiorystycznej¹¹. Powołując się na pracę brytyjskiego filozofa Johna Lucasa (*Minds, Machines and Gödel*), Hetmański stwierdza, że założenie o równoważności ludzkiego umysłu i programu komputerowego pod względem możliwości dokonywania operacji poznawczych wynika z przyjęcia, że

⁷ A. Turing, *On Computable Numbers with an Application to the Entscheidungsproblem*, ss. 230-265, w: "Proc. Lond. Math. Soc.", Nr 2(42)/1936 s. 231.

⁸ Por. A. Newell, H. Simon, *Computer simulation of human thinking and problem – solving*, w: *Management and the Computer of the Future*, red. M. Greenberger, Wiley, New York 1962, ss. 94-113.

⁹ Koncepcja Turinga jest bowiem formalnie równoważna teorii sformułowanej przez Alonzo Churcha. Zob. A. Church, *An unsolvable problem of elementary number theory*, w: "American Journal of Mathematics", Nr (58)/1936, ss. 345-363. Sam Turing następująco ujmuje związki swej koncepcji z teorią funkcji rekurencyjnych: „Opierając się na pewnych wynikach logiki matematycznej można wykazać, że istnieją granice możliwości maszyn dyskretnych. Najlepiej znanym z tych wyników jest twierdzenie Gödla (...). Istnieją inne, pod pewnymi względami podobne wyniki, które zawdzięczamy Churchowi (1936), Kleene'owi (1935) oraz Turingowi (1937). Ten ostatni wynik jest najwygodniejszy do rozpatrywania, ponieważ odnosi się bezpośrednio do maszyn, podczas gdy inne można stosować tylko w stosunkowo pośrednim argumentacie: np., gdybyśmy chcieli zastosować twierdzenie Gödla to musielibyśmy poza tym posiadać jakieś sposoby opisu systemów logicznych w terminach maszyn i maszyn w terminach systemów logicznych.” [A. Turing, *Maszyny liczące a inteligencja*, ss. 23-47, tłum. D. Gajkiewicz, w: *Maszyny...*, red. E. Feigenbaum, J. Feldman, ss. 33-34].

¹⁰ A. Turing, *On Computable...*, s. 250.

¹¹ Zob. M. Hetmański, *Umysł...*, s. 141.

plaszczyną porównywania zdolności poznawczych człowieka i maszyny mogą być jedynie zobiektywizowane, uzewnętrznione wypowiedzi obu fizycznych układów.¹²

Zatem, tylko na podstawie aktywności zewnętrznej maszyny (umysłu) orzeka się o jej/jego stanach mentalnych.

Łatwo zauważyć, iż zaakceptowanie przez Turinga behawiorystycznego podłoża swojej koncepcji implikuje szereg trudności teoretycznych. Abstrahując od ustaleń psychologicznych, J. Lucas argumentuje, że już porównanie człowieka z maszyną w zakresie przeprowadzania wnioskowania dedukcyjnego wystarcza, aby zaprzeczyć identyczności umysłu i maszyny pod względem dokonywania operacji poznawczych. Píše Lucas:

Dla każdej maszyny istnieje prawda, której nie można wytworzyć jako prawdziwej, którą jednak może wytworzyć człowiek. Pokazuje to, że maszyna nie może być pełnym i adekwatnym modelem umysłu. Nie może ona robić wszystkiego, co może zrobić umysł, ponieważ, mimo, że może ona wiele, to jednak pozostaje zawsze coś, czego nie może ona zrobić, a co może zrobić umysł. Nie znaczy to, że nie możemy zbudować maszyny do symulowania żadnego pożądanego umysłowego zachowania; znaczy to tylko, że nie możemy zbudować maszyny do symulowania każdego umysłowego zachowania. (...) Twierdzenie Gödla jest piętą Achillesową cybernetycznej maszyny. Tym samym nie możemy mieć nadziei na wytworzenie maszyny, która będzie w stanie robić wszystko, co może zrobić umysł; nie możemy nigdy mieć, nawet w zasadzie, mechanistycznego modelu umysłu.¹³

Wywód Lucasa oparty jest na argumencie głoszącym, iż **zaprogramowana maszyna cyfrowa nie może wykroczyć poza system formalny** (np. działać intuicyjnie), **ponieważ ona sama jest takim systemem**. Maszyna matematyczna jest więc zdeterminowana przez zbiór reguł formalnych, w oparciu który została zaprogramowana. Żaden bowiem program komputerowy – twierdzi Lucas – nie posiada zdolności dowodzenia prawdziwości zdań samozwrotnych, czyli takich, które wykraczają poza system.

Mogłaby ona tego dokonać – pisze M. Hetmański – lecz podczas czynności dowodzenia takiego zdania musiałaby dobudować dodatkową część swojego systemu; dla wykazania swej zupełności musiałaby zatem wykraczać poza siebie samą *ad infinitum*, co jest niemożliwe do przyjęcia. Byłaby wprawdzie coraz bogatszą maszyną, lecz czy inną maszyną, czy dorównywałaby umysłowi, pyta Lucas. Jego odpowiedź jest jednoznaczna – nie byłaby takim układem.¹⁴

Zdaniem Lucasa, twierdzenie Gödla pozwala wnosić, że tylko ludzki umysł jest w stanie rozpoznawać i korygować sprzeczności nie tylko w formułach matematycznych, ale przede wszystkim w swoim działaniu. Jedyne świadomy umysł

może zajmować się Gödłowskimi kwestiami w sposób, w który nie może maszyna, ponieważ świadoma istota może zarówno rozważać samą siebie, jak i swoje działania i to nie inaczej niż w tym działaniu. (...) Może (ona, przyp. P. K.) odzwierciedlać siebie samą i krytkować własne działania i do czego nie potrzebuje dodatkowej części; jest kompletna i nie ma żadnej pięty Achillesowej.¹⁵

¹² Tamże s. 158.

¹³ J. Lucas, *Minds, Machines and Gödel*, ss. 112-137, w: "Philosophy", vol. XXXVI, Nr (137)/1961, ss. 115-116.

¹⁴ M. Hetmański, *Umysł...*, s. 159.

¹⁵ J. Lucas, *Minds...*, s. 125.

Krytyka Lucasa, będąca w głównej mierze krytyką założeń leżących u podstaw testu Turinga (*gry w udawanie*), wydaje się być niepełna. Lucas zastosował bowiem strategię zbieżną z podejściem, które Turing nazwał *sprzeciwem matematycznym*¹⁶. Pierwszy z nich, sprzeciw opierający się na wnioskach płynących z twierdzenia Gödla, można odrzucić poprzez wskazanie, że dotychczas nie udowodniono, iż umysł ludzki nie podlega takim samym ograniczeniom, jakie przypisuje się maszynom cyfrowym. Innymi słowy,

krótką ripostą na ten argument, jest to, że chociaż ustalono, że istnieją granice możliwości każdej poszczególnej maszyny, to jednak jedynie bez dowodu stwierdzono, że żadne ograniczenia nie stosują się do ludzkiego intelektu.¹⁷

Zarzuty Lucasa nie odnoszą się jednak do kwestii związanych z introspekcyjnością i intencjonalnością stanów wewnętrznych maszyny cyfrowej. Autor *Minds, Machines and Gödel* nie analizuje więc koncepcji umysłu rozwijanych na gruncie nauk kognitywnych, przyjmujących za teoretyczną podstawę obliczeniowe ujęcie umysłu. Koncepcje te, o wyraźnie funkcjonalistycznym zabarwieniu wymagają, jak sądzę, namysłu, ze względu na oryginalność i wagę uzyskanych wyników. Z tej zatem przyczyny poniższa część pracy traktuje o funkcjonalistycznej teorii umysłu i jej związkach z programem badania sztucznej inteligencji.

3. Czym jest funkcjonalizm w filozofii umysłu?

Termin *funkcjonalizm* nieodłącznie wiąże się z postacią Hilarego Putnama. Filozof ten w ogłoszonym w 1960 r. artykule *Minds and Machines*¹⁸ pisał:

Maszyna będąca zdolna do identyfikacji przynajmniej jednego ze swych własnych stanów strukturalnych, znajduje się w położeniu analogicznym do sytuacji człowieka, który może odkryć niektóre, ale nie wszystkie dysfunkcje swego ciała ze zmiennym poziomem dokładności.¹⁹

Stwierdzenie to stało się kamieniem węgielnym funkcjonalistycznej teorii umysłu.

Podstawą twierdzenia Putnama o paralelach między maszyną cyfrową a człowiekiem jest logiczno-techniczna analiza maszyny Turinga. Maszyna ta, jako układ fizyczny, którego stany podlegają określonym prawidłowościom, nie jest, wg Putnama, tożsama ze swoimi stanami logicznymi. Maszyna Turinga może być bowiem zrealizowana na wiele sposobów, przy czym każda z tych realizacji może wykonywać te same operacje logiczne (np. obliczanie czy wnioskowanie).

Jeśli – stwierdza M. Hetmański – uniwersalna maszyna Turinga jest w stanie symulować (modelować) działanie każdej cyfrowej maszyny, to może się tak dziać właśnie ze względu na niezależność działania takiej maszyny na poziomie logicznym, (obliczeniowym) od działania na poziomie fizycznym. Maszyna Turinga może zatem znajdować się w takich dowolnych stanach, które nie zależą od rodzajów tej realizacji, konkluduje Putnam. Dla konkretnej maszyny na przykład 'bycie w stanie A' i

¹⁶ Zob. A. Turing, *Maszyny...*, s. 33.

¹⁷ Tamże, s. 34.

¹⁸ H. Putnam, *Minds and Machines*, w: *Dimensions of Mind. A Symposium*, red. S. Hook, New York 1961, ss. 138-164.

¹⁹ Tamże, s. 148.

'włączenie przełącznika 36' są różnymi jej stanami, każdy z nich dotyczy różnych jej własności – logicznych oraz fizycznych.²⁰

Główna teza funkcjonalizmu głosi więc, iż te same procesy poznawcze (np. percepcja, myślenie, wnioskowanie) mogą być zrealizowane przez różne układy fizyczne. Stąd też stany mentalne stanowią funkcję stanów fizycznych. Znaczy to, że identyfikacja danego stanu mentalnego, realizowanego przez system przetwarzający informacje, jest niezależna od fizycznej struktury tego systemu.

Identyfikacja ta – Pisze Putnam. – jest niezależna w tym sensie, że *wiara w to, że śnieg jest biały* stanowi ten sam stan obliczeniowy każdego fizycznie możliwego organizmu zdolnego do posiadania przekonań tego typu.²¹

Widocznym jest zatem, że procedury obliczeniowe determinują warunki spełniania danego stanu mentalnego. Opierając się bowiem na tezie Churcha-Turinga, można stwierdzić, iż dany stan mentalny jest identyczny z pewną konfiguracją realizującego go systemu²². Wynika to z założenia, że warunki spełniania procedur obliczeniowych (algorytmów) nie są wyznaczone przez żadną strukturę fizyczną. Dla przykładu: koniunkcja dwóch przesłanek jest fałszywa wtedy, gdy przynajmniej jedna z przesłanek jest fałszywa. Formuła ta jest spełniona niezależnie od tego, czy jest realizowana przez człowieka czy komputer. Zatem, konkluduje Putnam, wystąpienie danego stanu mentalnego jest warunkowane realizacją pewnego algorytmu przez system przetwarzający informacje.

Nietrudno dostrzec, że przedstawiony powyżej program badania umysłu jest co najmniej dyskusyjny. Kwestię tę podjął sam Putnam twierdząc, że jego stanowisko ogłoszone w *Minds and Machines* (tzw. *funkcjonalizm ortodoksyjny*) pociąga za sobą szereg trudności teoretycznych wynikających z twierdzenia Gödla. Twierdzenie o niezupełności systemów aksjomatycznych wskazuje, że nie jest możliwe jednoznaczne wyrażenie danego stanu mentalnego maszyny za pomocą skończonego zbioru reguł obliczeniowych²³. Nie sposób zatem ściśle wyznaczyć, czy te same stany funkcjonalne umysłu (np. wnioskowanie, obliczanie, percepcja) są realizowane przez tą samą procedurę co „stany mentalne” maszyny. Ponadto, ujęcie zaproponowane w ramach funkcjonalizmu skrajnego nie wyjaśnia pytań związanych z: problemem postaw propozycjonalnych, semantyczną interpretacją procedur obliczeniowych, czy intencjonalnym odniesieniem stanów mentalnych danego systemu. Podkreślić jednakże należy, iż w okresie funkcjonalizmu skrajnego, Putnam nie podejmuje próby eksplikacji tych zagadnień²⁴.

Aporie wynikające z twierdzenia Gödla, skłoniły Putnama do rewizji swej koncepcji. Funkcjonalizm skrajny został odrzucony na rzecz *funkcjonalizmu umiarkowanego*, którego podstawowe założenia zostały ogłoszone w pracy *Representation and Reality* z roku 1988. W niej to odrzucone zostało ontologiczne ujęcie funkcjonalizmu na rzecz epistemologicznej

²⁰ M. Hetmański, *Umysł...*, s. 71.

²¹ H. Putnam, *Why Functionalism Didn't Work*, ss. 441-459, w: tenże, *Words and Life*, Harvard University Press, Cambridge Mass. 1994, ss. 448-449.

²² Por. S. Bringjord, M. Zenzen, *Cognition Is Not Computation. The Argument From Irreversibility*, s. 285-320, w: "Synthese", Nr (113) / 1997, s. 311.

²³ Zob. H. Putnam, *Representation and Reality*, Cambridge Mass. 1988, s. XV.

²⁴ Por. S. Harnard, *Minds, Machines and Searle*, ss. 5-25, w: "Journal of Theoretical and Experimental Artificial Intelligence", Nr 1/1989, s. 13.

interpretacji tego stanowiska²⁵. Zmiana stanowiska teoretycznego Putnama wiąże się więc przede wszystkim z odrzuceniem skrajnego redukcjonizmu, który głosi, że stany funkcjonalne (np. myślenie) są tożsame ze stanami układu fizycznego. Co więcej, stwierdza autor *Representation...*, modelem umysłu nie może być maszyna Turinga, gdyż danego stanu funkcjonalnego nie można wyrazić za pomocą tylko jednej procedury formalnej. Należy również dodać, że (wg Putnama) analogia między umysłem ludzkim a maszyną Turinga byłaby zasadna wtedy i tylko wtedy, gdyby maszyna Turinga była tworem biologicznym, a w konsekwencji podlegałaby określonym prawom biologii²⁶.

W warstwie negatywnej, funkcjonalizm umiarkowany zaprzecza zatem determinacji działania określonego układu poznawczego (w tym przypadku – umysłu ludzkiego) przez zbiór formuł logicznych. Wydaje się, iż stwierdzenie to wynika z założenia przez Putnama niewspółmierności pomiędzy stanami obliczeniowymi a stanami funkcjonalnymi danego systemu. Stany obliczeniowe nie mogą być tożsame ze stanami funkcjonalnymi, ponieważ na poziomie opisu zachowania danego systemu pojawia się problem jednoznacznego przyporządkowania pewnego algorytmu danej reakcji systemu. Na przykład, każdy proces wnioskowania musiałby być, zgodnie z tym ujęciem, warunkowany realizacją danego algorytmu A. Jednakże, twierdzi Putnam

komputery badające różne analogie lub odbierające różne przyczynowe dane (...) mogą posiadać zupełnie odmienne opisy tych danych niż te, które zawierają formuły maszyny Turinga, czy jakiegokolwiek inne formuły. Mogą one posiadać różne numery stanów (obliczeniowych, przyp. P. K.) oraz działać na podstawie różnych reguł transformacji.²⁷

Z tego też powodu, utożsamienie stanów mentalnych i stanów obliczeniowych jest nieprawomocne.

Teza ta implikuje z kolei pozytywny wymiar umiarkowanego funkcjonalizmu Putnama. W *Representation and Reality* filozof stwierdza:

Porzuciłem funkcjonalizm (skrajny, przyp. P. K.), ponieważ sądzę, iż istnieją dowody potwierdzające fakt, że **stany umysłu są nie tylko ustrojowo plastyczne, ale również obliczeniowo plastyczne** (wytłuszczenie, P.K.). Rozumiem przez to, że fizycznie możliwe istoty, które sądzą, że w okolicy jest mnóstwo kotów, czy czegokolwiek innego, mogą mieć niezliczone najrozmaitsze «programy».²⁸

Przyjmując M. Hetmańskiego interpretację stanowiska Putnama²⁹, stwierdzić można, że funkcjonalne stany poznawcze składające się na umysł człowieka, nie są skończonym zbiorem formalnych reguł przetwarzania symboli.

Umysły – pisze Hetmański – nie są też programami, gdyż funkcjonalność stanów umysłowych pociąga za sobą ich zasadniczą wielość i zmienność, ich plastyczność. Każdy fizyczny układ poznawczy wyznacza (determinuje) za sprawą

²⁵ W sprawie tego rozróżnienia zob.: V. Harcastle, *Computationalism*, ss. 303-317, w: "Cognition", Nr 3/1995, ss. 312-313.

²⁶ Por. H. Putnam, *The mental life of some machines*, ss. 408-428, w: tenże, *Mind, Language and Reality*, Cambridge Mass. 1979, s. 412. Znamienym jest, iż artykuł ten, pisany jeszcze z pozycji funkcjonalizmu skrajnego zawiera przesłanki implikujące późniejsze odrzucenie tego stanowiska.

²⁷ H. Putnam, *Why Functionalism...*, s. 452.

²⁸ H. Putnam, *Representation...*, s. 139.

²⁹ Por. M. Hetmański, *Umysł...*, ss. 72-73.

wielokierunkowości i zmienności swojego działania wielość i różnorodność funkcjonalnych realizacji. Nie przybierają one jednej tylko postaci, ale jest ich wiele, w zasadzie, jak stwierdza Putnam, istnieje pod tym względem ich urodzaj i nadmiar.³⁰

Ważnym jest również to, że w ramach funkcjonalizmu umiarkowanego główna płaszczyzna rozważań oscyluje wokół ustalenia różnicy pomiędzy ludzką i komputerową realizacją stanów funkcjonalnych. Zarysowana przez Putnama teza o braku korelacji pomiędzy procesami obliczeniowymi dokonywanymi przez dany system a działaniem tego układu, prowadzi do wniosku, że tak zwane różnice *hardwarowe* (a więc fizyczne różnice w budowie ludzkiego umysłu i maszyny cyfrowej) określają działanie *software* (czyli treści sądów, przekonania, pragnień *etc.*) układu fizycznego. Godnym podkreślenia jest zatem, że

«poziom intencjonalny» nie daje się po prostu zredukować ani do «poziomu obliczeniowego» ani do «poziomu fizycznego»³¹

ze względu na problem postaw propozycjonalnych. Obliczeniowa plastyczność wyklucza bowiem ujęcie stanów funkcjonalnych jako układów monadycznych³² i (jeśli to założenie jest prawdziwe) prowadzi do stwierdzenia, że stany mentalne stanowią sieć wzajemnie implikujących się elementów, nie zaś izolowanych struktur warunkujących daną reakcję systemu. Jeżeli, na przykład, pewien system jest przekonany, że śnieg jest biały i zarazem ma nadzieję, że wieczorem skończy pisać artykuł, to koniunkcja tych stanów tworzy kolejny stan funkcjonalny. Z tej racji Putnam stwierdza, iż

funkcjonalizm obstający przy tezie, że postawy propozycjonalne są wyłącznie stanami obliczeniowymi (...) nie może być ujęciem poprawnym.³³

W świetle teoretycznego dysonansu pomiędzy funkcjonalizmem ortodoksyjnym i umiarkowanym, zastanawiającą jest przyczyna, dla której pierwszy z funkcjonalizmów Putnama stał się filozoficznym zapleczem dla teorii i praktyki Sztucznej Inteligencji. Mając na uwadze trudności funkcjonalizmu skrajnego oraz odrzucenie tego stanowiska przez samego autora *Representation...*, zdziwienie może budzić fakt, iż skrajny funkcjonalizm stał się dominującym kierunkiem w badaniach nad sztuczną inteligencją.

Byłem pierwszym filozofem – wspomina Putnam – który postawił tezę, że komputer jest właściwym modelem umysłu. Doktrynę tą określiłem mianem *funkcjonalizmu*. Stała się ona wiodącym punktem widzenia – niektórzy sądzą nawet, że ortodoksją – we współczesnej filozofii umysłu.³⁴

Abstrahując jednak od kontekstu odkrycia funkcjonalizmu skrajnego dla teorii i praktyki AI, warto podjąć namysł nad rolą funkcjonalizmu w badaniach nad sztuczną inteligencją i prześledzić konsekwencje tego stanowiska dla programu badawczego AI.

³⁰ Tamże, ss. 72-73.

³¹ H. Putnam, *Representation...*, s. 140. Zob. także H. Putnam, *The nature of mental states*, w: tenże, *Mind...*, ss. 429-440.

³² Stan funkcjonalny jest monadyczny wtedy, gdy jego struktura wyznaczana przez inną relację niż postawa propozycjonalna (np. przekonanie) – znaczenie – postawa propozycjonalna2 (np. odczucie). Por. J. Fodor, *Jak grać...*, ss. 29-35.

³³ H. Putnam, *Why functionalism...*, s. 444.

³⁴ H. Putnam, *Representation...*, s. XI

4. Funkcjonalizm w badaniach nad sztuczną inteligencją

W badaniach nad sztuczną inteligencją, stanowisko funkcjonalistyczne przybiera nazwę *funkcjonalizmu komputerowego*³⁵. Funkcjonalizm komputerowy od funkcjonalizmu ortodoksyjnego Putnama odróżnia jedynie przedmiot badań. O ile funkcjonalizm skrajny odnosi się do wszystkich procesów poznawczych umysłu/maszyny, to funkcjonalizm komputerowy bada tylko jeden aspekt problemu – mianowicie ludzką/maszynową inteligencję. Należy przy tym zaznaczyć, że funkcjonalizm komputerowy stosuje tę samą aparaturę pojęciową, co funkcjonalizm ortodoksyjny.

Zbieżności pomiędzy funkcjonalizmem komputerowym a funkcjonalizmem skrajnym, dają się zauważyć już u początków kształtowania się teorii Sztucznej Inteligencji. Marvin Minsky pisał bowiem w roku 1963:

Wobec tego «czym właściwie jest» inteligencja? Z mojego punktu widzenia jest to raczej zagadnienie estetyki albo szacunku, niż nauki. Dla mnie «inteligencja» oznacza niewiele więcej niż kompleks działań, z którym mamy do czynienia, ale którego nie rozumiemy. Podobnie sprawa wygląda z zagadnieniem «głębi» w matematyce. Gdy tylko dowód twierdzenia jest zrozumiały to jego treść wydaje się trywialna.³⁶

Już na podstawie powyższej wypowiedzi Minsky'ego nietrudno dostrzec analogie pomiędzy funkcjonalizmem komputerowym a funkcjonalizmem skrajnym. W funkcjonalizmie komputerowym inteligencja nie stanowi bowiem koniecznej i dostatecznej racji odróżniającej umysł ludzki od maszyny. **Inteligencja jest raczej ogólną własnością wszystkich systemów poznawczych wykazujących w danych sytuacjach zdolność do rozwiązywania problemów.**

W odniesieniu do maszyn cyfrowych – stwierdza M. Hetmański – polega to na wykonywaniu programu działania, zaś w przypadku ludzi – na realizacji ich działań ze względu na pewien zbiór reguł czy zasad. Inaczej mówiąc, domniemana inteligencja maszyn wykazuje istotne podobieństwa do takiego działania człowieka, przy którego spełnianiu normalnie orzeka się, że jest ono inteligentne lub, inaczej mówiąc, wymaga inteligencji³⁷.

Konsekwencją akceptacji powyższych przesłanek jest przyjęcie funkcjonalistycznej (operacyjnej) definicji inteligencji w badaniach AI. W sformułowaniu Herberta Simona definicja ta głosi:

Inteligencja nie jest sprawą substancji – protoplazmy, szkła czy drutu – lecz formy, którą substancja przyjmuje i procesów, które w niej zachodzą. Korzeniami inteligencji są symbole z ich denotacyjną mocą i podatnością na manipulowanie. Symbole mogą zaś być wytwarzane niemalże ze wszystkiego, co może zostać zgromadzone, połączone i zorganizowane. **Inteligencją jest umysł implementowany w każdy wymodelowany rodzaj materii (wytluszczenie, P. K.).**³⁸

³⁵ Zwolennikami podejścia funkcjonalistycznego są niemal wszyscy zwolennicy tzw. mocnej wersji Sztucznej Inteligencji [*strong AI*]. Tytułem przykładu wymienić należy koryfeuszy badań nad sztuczną inteligencją: A. Newella, H. Simona, M. Minsky'ego i D. Hofstadtera. W kwestii rozróżnienia rodzajów Sztucznej Inteligencji zob. J. Searle, *Umysł, mózg i nauka*, tłum. J. Bobryk, Warszawa 1995, ss. 25-28.

³⁶ M. Minsky, *Na drodze do stworzenia sztucznej inteligencji*, ss. 378-428, w: *Maszyny matematyczne i myślenie*, red. E. Feigenbaum, J. Feldman, tłum. P. Kerntopf, Warszawa 1972, s. 420.

³⁷ M. Hetmański, *Umysł...* s. 51.

³⁸ H. Simon, *The Science of the Artificial*, Cambridge Mass., London 1996, s. 35.

W teorii i praktyce Sztucznej Inteligencji nie jest zatem istotne pytanie o sposób istnienia i genezę umysłu. Komputerowy funkcjonalizm nie podnosi również problematyki oscylującej wokół zagadnienia świadomości, problemu dualizmu psychofizycznego, czy tzw. kwestii dostępności do innych umysłów [*other-minds problem*]. Stanowisko to skupia się raczej na próbie udzielenia odpowiedzi na pytanie o konieczne i konstytutywne warunki możliwości zachodzenia określonych procesów poznawczych w maszynie/umyśle ludzkim, a także o sposób takiego modelowania (programowania) maszyn cyfrowych, aby ich działanie na poziomie behawioralnym było nieodróżnialne od działania człowieka. Refleksja ta nie ma jednak charakteru ontologicznego. Badanie inteligencji ma raczej metodologiczny status. Polega ono na poszukiwaniu procedur, za pomocą których możliwa stanie się konstrukcja systemów działających w sposób nieodróżnialny/porównywalny z działaniem człowieka³⁹.

Poglądy badaczy inteligencji na sposób istnienia inteligencji ludzkiej/maszynowej w głównych punktach programu są więc zgodne. Główne tezy przedstawić można następująco:

1. Inteligencja nie ma charakteru substancjalnego. Jest ona funkcją maszyn/umysłów działających w oparciu o procedury obliczeniowe (zbiór aksjomatów i reguł logiki formalnej).
2. Stąd też, o istocie ludzkiej/maszynowej inteligencji nie decyduje materia, z której dany układ został wykonany, lecz efekt działania tego układu.
3. Jeżeli inteligencja jest warunkiem istnienia procesów poznawczych umysłu/maszyny, to dla realizacji tych procesów rodzaj podłoża fizycznego, na którym są realizowane, jest równie marginalny, jak w przypadku inteligencji.

W świetle powyższych uwag, sądzić można, że określając swe stanowisko teoretyczne, funkcjonalizm komputerowy w takiej samej mierze bazował na tezach ortodoksyjnego funkcjonalizmu Putnama i rozstrzygnięciach teorii behawiorystycznych. Teza o rozpoznawaniu inteligencji umysłu/maszyny poprzez badanie efektu działania układu nasuwa skojarzenia nie tylko o inspiracjach twierdzeniami metodologicznego behawioryzmu Watsona, ale przede wszystkim – tezami behawioryzmu logicznego Ryle'a i Quine'a. Wydaje się bowiem, iż postulat komputerowego funkcjonalizmu mówiący, że o tym, czy maszyna/umysł wykazuje inteligencję, decydują efekty działania układu, jest tożsamy z postulatem dyspozycji mentalnych, głoszonym przez logicznych behawiorystów. W tym rozumieniu, inteligencja stanowi dyspozycję (funkcję) wyspecjalizowanych układów fizycznych. Określenie jej istoty (substancji) nie jest więc możliwe. O jej istnieniu przesądza sposób zachowania się układu w sytuacjach problemowych.

³⁹ Por. M. Minsky, *Minds Are Simply What Brain Do*, <http://www.leaderu.com/softcopy/truth/2truth03.txt>.