prof. dr hab. Marek Hetmański
Instytut Filozofii
UMCS

Lublin, 30 maja 2022 r.

Recenzja rozprawy doktorskiej mgr Łukasza Sarowskiego
pt. „Robot humanoidalny jako podmiot życia społecznego.
Problemy filozoficzne i społeczne"

Rozprawa doktorska została napisana pod kierunkiem naukowym dr hab. Małgorzaty Gruchoły prof. KUL na seminarium Metodologii nauk o kulturze. Ma charakter interdyscyplinarny, gdyż łączy analizy podmiotowości poznawczej humanoidalnych robotów (w tym znaczeniu porusza się w dyscyplinie filozofia, ale także w kognitywistyce) z analizami społecznych interakcji takich robotów z człowiekiem (nawiązując do koncepcji socjologicznych, chociaż w stopniu o wiele mniejszym). Odwołuje się także do rozważań z antropologii kulturowej, jak również do koncepcji funkcjonujących w informatyce i robotyce. Z każdej z tych dyscyplin Autor rozprawy przejmuje terminologię i definicje oraz klasyfikacje oraz stanowiska i koncepcje, starając się analizować problem autonomii poznawczej humanoidalnych robotów z uwzględnieniem jego wieloaspektowości oraz wątków pobocznych, takich jak sprawstwo czy odpowiedzialność sztucznych podmiotów. Głównym celem rozprawy jest, jak stwierdza, opracowanie aparatury pojęciowej dla jednolitego sformułowania tytułowego zagadnienia oraz wyliczenie, a następnie scharakteryzowanie głównych, konstytutywnych własności sztucznych podmiotów poznawczych. W tym sensie rozprawa ma charakter referujący stan wiedzy w odniesieniu do tytułowego zagadnienia oraz rekonstruujący problem podmiotowości sztucznych podmiotów od strony filozoficznej, głównie epistemologicznej.

Wartość merytoryczna rozprawy

Teza rozprawy jest wyraźnie sformułowana – rozwój technologii informatycznych i robotyki, przyczyniając się do stworzenia coraz bardziej złożonych i samodzielnych urządzeń i robotów przyczynia się do zmiany reakcji i postaw ludzi wobec nich, co wyraża się w przydawaniu im autonomii działania oraz statusu podmiotów życia społecznego.

Autor rozprawy wskazuje na dokonania w dziedzinie sztucznej inteligencji i robotyki, w ramach których powstają generacje autonomicznych maszyn. Przytacza koncepcje i teorie mówiące o autonomii maszyn, które powstają od paru dekad w ramach tzw. filozofii robotyki, przywołuje także koncepcje socjologiczne mówiące o specyfice społecznych relacji. Na tej podstawie stara się ocenić wartość poznawczą tezy mówiącej o wytwarzaniu się wyobrażeń o nowej autonomii podmiotowej i społecznej humanoidalnych maszyn i robotów. Jak sam pisze: "Przedmiotem moich zainteresowań jest sposób rozumienia robota humanoidalnego jako podmiotu życia społecznego na gruncie filozofii robotyki" (s. 5). Mówiąc dokładniej, Autor analizuje i ocenia nastawienia i wyobrażenia, jakie ludzie (głównie teoretycy z dziedziny filozofii robotyki) mają na temat rosnącej autonomii funkcjonowania maszyn. Humanoidalny robot traktowany jest jako artefakt technologiczny – obiekt i zjawisko osadzone w życiu społecznym wytworzonym przez współczesną technologię komputerową. Aby argumentować na rzecz podmiotowości społecznej robotów, Autor próbuje

doprecyzować, jak pisze, „pojęcie podmiotu życia społecznego, które może być aplikowane w badaniach nad robotyką humanoidalną", co stara się osiągnąć poprzez wpisanie swoich analiz w „paradygmat myślenia o technologii w kontekście przyszłych relacji społecznych" (s. 7). Jest to, trzeba przyznać, zadanie ambitne, być może nawet za szerokie jak na standardową rozprawę doktorską, świadczące niemniej, że Autor widzi analizowany problem w szerokim – cywilizacyjnym i kulturowym aspekcie. W rzeczywistości rozprawa doktorska jest analizą pojęć, jakie łączą się z podmiotowością i autonomią działań człowieka w ich zestawieniu i porównaniu z podmiotowością maszyn, a także powiązanych z nimi pojęć cielesności, komunikowalności oraz interakcji społecznych. Z tego powodu Autor pisze, że jego dysertacja „ma charakter metaprzedmiotowy z ambicjami merytorycznymi" i zmierza do „wskazania własności służących analizie i argumentacji na rzecz podmiotowości społecznej robotów humanoidalnych", dokładniej mówiąc, „przyznania lub nieprzyznania statusu podmiotu życia społecznego robotom" (s.10-11). Na tej podstawie można stwierdzić, że rozprawa doktorska mgr Sarowskiego porusza znaczący merytorycznie oraz ważny filozoficznie problem istoty podmiotowości działań poznawczo-praktycznych człowieka uwikłanych w rosnącą złożoność i częściową autonomię humanoidalnych robotów.

Tematyka rozprawy i jej charakterystyka

Rozdział pierwszy, który ma nieprecyzyjny tytuł „Założenia terminologiczne", sugerujący zaledwie analizy językoznawcze, jest w rzeczywistości poświęcony wyliczeniu i charakterystyce cech podmiotowości jako takiej poprzez ich zestawienie w dwóch ujęciach: antropocentrycznym i nieantropocentrycznym. Rozpoczyna się od znaczenia terminu „subiectum" stosowanego tradycyjnie do człowieka, co jest omówione poprzez odwołanie się do stanowisk I. Kanta i E. Husserla, a także K. Wojtyły i M.A. Krąpca. Po czym następują analizy pojęć autonomii i automatyzmu jako nietożsamych z podmiotowością ludzką. Wyróżnienie dwóch typów automatyzmu – mimowolności i powtarzalności działań ludzkiego ciała – pozwala Autorowi na podjęcie tematu autonomii ludzkiej podmiotowości, w tym takich jej cech dystynktywnych, jak świadomość i rozumność. Jest to dokonane poprzez odwołanie się do klasyfikacji i analiz pojęciowych Adama Węgrzeckiego, ale także innych badaczy – Alaina Touraine'a, Józefa Kozieleckiego, w szczególności Georga Simmela, od którego przejęte zostają pojęcia ja indywidualnego, ja uogólnionego oraz ja ogólnego jako użyteczne dla modelu społecznych interakcji. Dużo miejsca w rozdziale zajmują analizy cech podmiotowości formułowanych, jak pisze Autor, w „ujęciu nieantropocentrycznym", co lepiej byłoby ująć jako ujęcie pozaludzkie czy pozabiologiczne. Jest ono rozpatrzone w ramach tzw. humanistyki nieantropocentrycznej, która bardzo szeroko mówi o różnych „konfiguracjach podmiotowości" – zwierząt, klonów, cyborgów, a nawet samych rzeczy – co Autor podejmuje w swoich dalszych analizach. Omówiona zostaje również koncepcja aktora sieci Bruno Latoura, w której linia podziału między ludźmi, zwierzętami, maszynami i rzeczami jest zmienna, konwencjonalna, a nawet (jak twierdzi Latour) negocjowana, zaś aktor (aktant), czyli artefakt jest sumą relacji, w które wchodzi on w danych sytuacjach praktyczno-poznawczych.

W środkowej części rozdziału pierwszego mgr Sarowski omawia szczegółowo dyskusje, które toczą się w różnorodnych dyscyplinach nad podmiotowością humanoidalnych robotów, zwłaszcza ich społecznej podmiotowości i sprawstwa, jakim się one charakteryzują. Podkreśla wielokrotnie, że chodzi tutaj o podmiotowość i sprawstwo, które jest przypisywane takim robotom. Określenie „przypisywane" jest wielokrotnie używane dla podkreślenia konwencjonalności tych operacji interpretacyjnych. Autor mówi o „wrażeniu posiadania przez obiekt danych własności", także o „przypisywaniu życia obiektom nieożywionym" (s. 31), stawiając ważkie pytanie: „skąd się biorą w człowieku skłonności antropomorfizacyjne,

skutkujące ożywieniem rozmaitych obiektów?". Pytając o przyczyny owej „animacyjności", zauważa, że może ona wynikać z tendencji tzw. kultur prymitywnych, jak również z tendencji do antropomorfizowania spotykanego u dzieci i osób z zaburzeniami psychicznymi, spotykanego również magicznym myśleniu.

Dużo miejsca poświęca się w tej części rozdziału opisowi interakcji człowieka z robotem w ramach – mniej lub bardziej autonomicznego i świadomego (ze strony człowieka) – ich współdziałania, które opisywane jest nie tylko od strony fizycznej (technicznej), lecz także, jak pisze Autor, „symbolicznej, np. w oparciu o interfejsy werbalne i pozawerbalne, gdzie tworzona jest symbiotyczna przestrzeń interakcji między człowiekiem i robotem" (s. 33). Wymiar symboliczny określony jako obustronne korzyści dla człowieka i robota humanoidalnego wydaje się jednak nazbyt optymistyczną wizją, gdyż zadekretowana, lecz enigmatycznie określona „swoista współpraca" nie jest przez Autora w dalszej części rozdziału dookreślona. Przywołanie (za Moraną Alać) koncepcji ucieleśnienia przeżyć i doświadczeń robotów niewiele tutaj wyjaśnia, podobnie jak koncepcja gestów wykonywanych (imitowanych raczej) przez same roboty oraz ludzkiej na nie reakcji. Obie koncepcje nie mówią o symbiozie między człowiekiem i robotem, ani tym bardziej nie wskazują na symboliczny charakter takich relacji; symbolicznego poziomu, wraz z jego znakowym reprezentowaniem cech otoczenia i działania, wcale tutaj nie ma. Jeśli nawet Autor przytacza (za Sherry Turkle) przykłady emocjonalnych reakcji ludzi na imitowane przez roboty gesty, to reakcje takie nie dowodzą bynajmniej symbolicznego charakteru „gestu" robota, a co najwyżej prostego praktycznego znaczenia przypisanego mu przez reakcję człowieka. W obu przypadkach nie ma, tak ważnej w semiotyce, intencji znaczeniowej, jest zaledwie reakcja emocjonalna; nie zachodzi zatem zjawisko semiozy z intencjonalnością człowieka.

Sporo miejsca mgr Sarowski poświęca w tym rozdziale analizie pojęcia robota, sięgając do definicji sformułowanych w naukach technicznych, zwłaszcza w cybernetyce, na podstawie których formułuje cztery kryteria autonomiczności robotów – (1) sterowanie w układzie zdeterminowanym; (2) adaptację w oparciu o percepcję otoczenia; (3) stosowanie strategii działania oraz (4) podejmowanie samodzielnych decyzji. W nawiązaniu do nich przytacza także definicję (za Duffym i jego komentatorami) humanoidalnego robota jako złożonego urządzenia technicznego zaprojektowanego w celu wywoływania społecznych interakcji za pomocą imitujących człowieka oraz (nawiązujących do ludzkiego ciała) kształtów, gestów, reakcji, wobec których reagujący na nie użytkownik takiego urządzenia przyjmuje postawę tzw. społecznego zachowania. Mgr Sarowski pisze o „przejawianiu tendencji postrzegania robotów jako aktorów społecznych" (s. 38) – tendencji charakteryzującej tak użytkowników, jak i badaczy tych interakcji. Zostaje to na ogół uznane jako wystarczający wyróżnik społecznych działań między robotami a ludźmi. Za kolejny wyróżnik autonomii robotów uznawane jest ich „ucieleśnienie" na poziomie poznawczym, a także „komunikacyjność" na poziomie interakcji, o których również sądzi się, że odbywają się w „sferze symbolicznej i emocjonalnej", czego jednak Autor (referujący te wyróżniki za innymi badaczami) ani nie precyzuje, ani nie ocenia co do ich ważności.

W podsumowaniu rozdziału o autonomii humanoidalnych robotów mówi się o szeroko rozumianych kompetencjach: instrumentalnych, kognitywnych, aksjologicznych i refleksyjnych, które łącznie mają stanowić o podmiotowości człowieka, w tym również humanoidalnego robota. Autor rozprawy uważa, że kompetencje takie mogą być implementowane w budowie i działaniu robotów. Wyraża to przekonanie w metaforycznym zdaniu: „Kompetencje te znajdują zakotwiczenie we własnościach robotów, a więc ich ucieleśnieniu, autonomiczności i komunikacyjności" (s. 42). Trudno uznać sens tego sformułowania za trafny, gdyż zależność cech robotów od cech i własności działania człowieka (jego kompetencji stanowiących o podmiotowości, jak chce Autor) należałoby ująć

odwrotnie. To kompetencje robotów powinny być „zakotwiczone" (jeśli użyć tej metafory dla „ucieleśniania", czyli zrealizowania technicznego) w kompetencjach człowieka, a nie odwrotnie; wszak cechy budowy i funkcjonowania robotów są najpierw modelowane, a następnie implementowane na podstawie budowy ciała człowieka i jego określonego działania. Autor zauważa niemniej, że w ramach robotyki i w koncepcjach podmiotowości przypisywanej humanoidalnym robotom (w tzw. filozofii robotyki) dominuje interakcyjne pojmowanie podmiotowości społecznej tychże robotów. Wynika to z domniemania, że istnieje coś takiego jak społeczno-kulturowy proces doświadczania robotów, określony w ostatnim akapicie rozdziału jako „fizycznie usytuowane".

Drugi rozdział zatytułowany „Humanoidalne ucieleśnienie robotów" poświęcony jest w pierwszej części omówieniu badań w filozofii, psychologii i kognitywistyce nad rolą ciała w poznawaniu i tworzeniu świadomości, dlatego mówi się o „poznawczym ucieleśnieniu". Krótko scharakteryzowane są koncepcje ciała w fenomenologii E. Husserla i M. Merleau-Ponty'ego, bardziej zaś szczegółowo koncepcje cielesności takie jak: minimalne ucieleśnienie, ucieleśnienie biologiczne, ucieleśniona semantyka, funkcjonalizm, wreszcie nieco szerzej enaktywizm i stanowisko F. Vareli, A. Noe'go i E. Thompsona, nawiązujące również do fenomenologii, w których mówi się o czasowym wymiarze doświadczenia cielesnego. Przytoczone jest stanowisko M. Wilson i jej sześć ujęć cielesności poznania ludzkiego, a także koncepcja ucieleśnienia T. Ziemkego, przywołane są również uwagi H. Dreyfusa o „odcieleśnieniu" syntetycznych sieci neuronowych. Na podstawie tych koncepcji i wypracowanych w ich znaczeń ucieleśnienia Autor stara się stworzyć spójne rozumienie cielesności poznawczej, która ma być punktem odniesienia (modelem, wzorem do implementacji) dla robotów humanoidalnych. Jej cechą podstawową jest funkcjonowanie organizmu (ciała) w otoczeniu, w konkretnym środowisku percypowanym, a następnie umysłowo przedstawianym, reprezentowanym i modelowanym. Analizując zależność percepcji od cielesności w działaniu ludzkiego podmiotu, mgr Sarowski stwierdza słusznie, że rodzaj ucieleśnienia determinuje sposób spostrzegania świata. Konkluduje niemniej w sposób nie do końca zrozumiały, ani też wystarczająco uzasadniony i uargumentowany: „Zatem procesy poznawcze robotów humanoidalnych powinny tylko do pewnego stopnia odpowiadać procesom poznawczym człowieka" (s. 56). Nie mówi jednak, jaki rodzaj „powinności" miałby decydować o stopniu „nieodpowiedniości" między percepcją ucieleśnioną człowieka a percepcją robota. Bez doprecyzowania tej kwestii przechodzi następnie do omówienia kwestii „wymiaru społecznego ucieleśnienia robotów".

Kwestia ta – kluczowa dla głównego problemu rozprawy doktorskiej i jej tezy – jest sformułowana nieprecyzyjnie od strony użytych określeń, nie jest też do końca zrozumiała co do sformułowanych wniosków i tez. W paragrafie 2.3. Autor pisze, że „humanoidalność, zwłaszcza w perspektywie interakcji społecznych, nie zasadza się wyłącznie na funkcjach poznawczych robota", lecz „na odbiorze" cielesności robota przez człowieka wchodzącego z nim w interakcje, a także na fakcie, że „pozwala [to ucieleśnienie – M.H.] wiązać z robotami określone przekonania dotyczące ich realności oraz sprawczości w kontekście społecznych interakcji" (s. 56). Niejasny jest w tym stwierdzeniu nie tyle „odbiór" przez człowieka cielesności – o tym Autor rozprawy już wcześniej trafnie pisał, że chodzi o interpretację zachowań robotów przez ich użytkowników – lecz sformułowanie mówiące o „wiązaniu przekonań (...) w kontekście...". Jest ono nieprecyzyjne i przez to mylące, gdyż nie wiadomo, czy: (1) człowiek przypisuje robotom „określone przekonania" jako ich autonomiczne kompetencje poznawcze, co byłoby bardzo mocną i kontrowersyjną tezą, czy też: (2) chodzi tylko o to, że można o domniemanej społecznej interakcji i społecznym zachowaniu robotów żywić oraz wypowiadać (w tym sensie „wiązać") jakieś opinie, sądy i tezy w ramach dyskutowanej w rozprawie koncepcji społecznych interakcji z robotami (a więc „w kontekście"), co byłoby tezą słabszą, dającą się obronić. Z sensu przytoczonego

powyżej zdania nie wiadomo, która z możliwości wchodzi w rachubę i w jakim znaczeniu robot miałby (mógłby) mieć jakieś przekonania.

Paruzdaniowy akapit otwierający paragraf niniejszego rozdziału zatytułowany „Wymiar społeczny ucieleśnienia robotów" jest zbyt skrótowy i niejasny w sformułowaniach, aby można było do końca wywnioskować, którą z dwóch opcji Autor przyjmuje i za którą się opowiada, a jest to kwestia ważna dla całej tezy. Niejasne jest także używanie normatywnego sformułowania „powinny" w odniesieniu do trzech rodzajów ucieleśnienia (fizyczności, reakcji, „rozumienia świata kultury"), które miałoby być odniesione do projektowania i budowy robotów. Ta założona powinność (możliwość) – zapewne nie etyczna, lecz zaledwie techniczna – wyrasta z faktyczności i konieczności, jakie stwarza robotyka, czego jednak w rozprawie nie omawia się dokładniej. W sumie tzw. hipoteza podobieństwa, mówiąca, że posiadanie przez robot humanoidalny któregoś z rodzaju ucieleśnienia jest warunkiem traktowania go jako partnera w interakcjach z człowiekiem nie jest w rozprawie dobrze uzasadniona. Autor wielokrotnie ukazuje ważne jej aspekty – np. że „skala owego podobieństwa determinuje skłonność człowieka do użycia siebie jako kryterium" – rzadko jednak rozwija takie uwagi, pozostawiając je rozproszone, bez jakiejś syntezy. Końcowa uwaga mówiąca, iż „elementem wywołującym wrażenie realności jest określony rodzaj zachowania się robota, zgodny z przyjętymi normami i zasadami obowiązującymi w społeczeństwie" jest trafnym postawieniem problemu badawczego – społeczne zasady współdziałania wyznaczają konceptualizowanie i rozumienie domniemanych „społecznych" zachowań robotów – ale również nie znajduje rozwinięcia i filozoficznego podsumowania. Trafna uwaga Herberta Simona na ten temat (o mocach poznawczych robotów decydują wyłącznie warunki techniczne, a nie ich sztuczny intelekt) także nie znajduje rozwinięcia.

Rozdział trzeci poświęcony jest "komunikacyjnej naturze podmiotowości społecznej robotów humanoidalnych", którą Łukasz Sarowski rozpatruje przede wszystkim od strony jej dialogiczności. Wykorzystuje koncepcję dialogu i swobodnej rozmowy, bliskości z Innym sformułowaną przez Martina Bubera i Emanuela Levinasa, zwłaszcza zawartą w niej koncepcję dialogowej struktury osobowości. Na jej tle stawia pytanie o możliwość dialogu człowieka z robotem pozbawionym świadomości i refleksyjności. Pyta także o możliwość osobowości robotów jako cechy (ich wyposażenia, ucieleśnienia) warunkującej dialogiczny charakter relacji człowiek–robot. Tę domniemaną osobowość humanoidalnego robota – jego „osobowość społeczną" – traktuje jako efekt, jak pisze, „projektowania określonych cech społecznych na zasadzie wskazanego podobieństwa„ (s. 67), a także „obdarzając go pewnego rodzaju podmiotowym charakterem" (s. 68). Tutaj znowu Autor wskazuje na intencjonalny charakter przypisywania (dokładniej – projektowania) ludzkich cech robotowi, co ma być, jak się zakłada, podstawą dysponowania przezeń osobowością. Czy konieczną, czy też tylko możliwą (jedną z wielu) – tego nie dopowiada. Tym samym związek domniemanej osobowości robota z komunikacyjnym, w tym dialogicznym (równie domniemanym) charakterem relacji człowieka z robotem, nie zostaje należycie pogłębiony.

Sporo miejsca poświęcone jest analizie związku komunikacyjności robotów z interaktywnością w ich funkcjonowaniu; tę drugą cechę Autor rozprawy traktuje jako uproszczoną formę komunikacji „na gruncie robotyki społecznej". Analizowane są także interfejsy jako rodzaj współpracy człowieka z robotem. Jako ważny problem teoretyczny i badawczy zostaje omówiona kwestia werbalnej interaktywności w relacjach człowiek–humanoidalny robot. Autor formułuje (idąc za innymi badaczami) trzy warunki skutecznej komunikacji werbalnej robot – człowiek: (1) musi ona być osadzone w znanym robotowi środowisku; (2) być otwarta i nieokreślona; (3) wykorzystywać mowę – uznając je za trudne do realizacji w praktyce ze względu na wieloznaczność mowy ludzkiej, która jest wciąż poza zasięgiem rozpoznania i przetwarzania przez roboty. Odnosi się także do koncepcji socjologicznej Jonathana Turnera, który mówi o złożonym kontekście, w jakim przebiegają

społeczne interakcje, którego elementy (np. treści kultury czy ekologii, zwłaszcza emocje) są nieosiągalne w przypadku domniemanych społecznych zachowań humanoidalnych robotów. O trudnościach takich Autor mówi skrótowo, że są one „semantyczno-techniczne".

W rozdziale ostatnim „Autonomiczność działań robotów humanoidalnych" podjęta zostaje kluczowa dla całej rozprawy doktorskiej kwestia rodzaju i stopnia swobody działania robotów. Autonomia rozumiana jest jako niezależność (niezawisłość) działania podmiotu od czynników zewnętrznych (w tym innych podmiotów), dokonywania zmian w otoczeniu, jak również braniu odpowiedzialności podmiotu za działanie i zmiany nim wywołane. Autonomię działania podmiotu Autor rozpatruje w wymiarze antropocentrycznym oraz, co dla rozprawy najważniejsze, nieantropocentrycznym, czyli w odniesieniu do robotów humanoidalnych; oba rodzaje tytułowej kategorii i płaszczyzny jej analizy stale ze sobą zestawia i porównuje, co w niektórych przypadkach (np. analiz wolnej woli czy prawnych skutków podmiotowości i jej zróżnicowanej autonomii) jest niejasne i przez to mylące. W referowaniu tego tematu Autor niejednokrotnie (jak i w innych rozdziałach) miesza ze sobą porządek opisu tego zjawiska z jego normatywnością, czego przykładem jest zdanie: „Zaprezentowane powyżej rozumienie autonomiczności adaptuje się w obszar zagadnień nad sztucznymi systemami poznawczymi, które muszą potrafić dostosowywać się do zmiennych warunków otoczenia" (s. 89); zły styl zdania jeszcze bardziej wzmacnia pomieszanie obu porządków i nie wiadomo dokładnie dlaczego roboty „muszą" dostosowywać się do otoczenia.

Rozpatrywana jest kwestia odpowiedzialności za działanie człowieka i humanoidalnego robota, która jest, zdaniem mgr Sarowskiego, „dzielona pomiędzy człowiekiem a maszyną" (s. 91), a nawet skalowalna ze względu na stopień uczestnictwa programu czy maszyny w interakcji. Dlatego też mówi się dalej o podejmowaniu decyzji przez roboty, wraz z ich autonomią, kontrolą i odpowiedzialnością. Kwestię odpowiedzialności robotów Autor rozpatruje w ramach definicji i koncepcji tzw. etyki robotów („roboetyki"), działu etyki stosowanej, która odnosi się do tradycyjnych problemów i stanowisk etyki ogólnej. Wymienia szereg kwestii, które rodzą się w ramach tej etyki, lecz zaledwie je zaznacza; pokazuje, że wynikają one z coraz większego udziału techniki i maszyn w życiu człowieka, poprzestając na ukazaniu ich złożoności i otwartości. Także i tutaj opis tego zjawiska jest zestawiany i mieszany z normatywnymi określeniami w stosunku do przytaczanych i opisywanych faktów, czego przykładem jest zdanie: „(...) maszyny spełniające określenia podmiotowe mogą aspirować również do 'awansu społecznego' poprzez zaprzestanie postrzegania ich w kategoriach wyłącznie instrumentalnych" (s. 96). Zawiły styl zdania, w tym niejasność określenia o jakowymś „awansie" robotów do poziomu etycznych podmiotów, wzmacnia jeszcze bardziej błąd nieuprawnionego przejścia z poziomu opisu do poziomu normy; nie wiadomo ponadto, kto/co miałby zaprzestać „postrzegania" robotów jako tylko maszyn.

Co pewien czas Autor jednak stawia właściwe pytania o autonomię decyzji robotów, zwłaszcza kiedy odwołuje się do innych autorów, jak na przykład do ważnego stanowiska J.P. Sullinsa, który pokazuje odmienność sytuacji działania robotów – ze względu na rolę społeczną, jaką pełnić może robot – na tle dylematów odpowiedzialności sprawczej podmiotów ludzkich. W tym kontekście przytoczone zostają tzw. prawa robotów Isaaca Asimowa i dyskusja, jaką ten literacki projekt wywołał w filozofii i etyce. Wskazana zostaje także interesująca kwestia ingerencji człowieka w funkcjonowanie robota (ściślej sprzężonego układu: człowiek – robot) jako kryterium stopnia autonomii decyzji humanoidalnego robota, a także stopnia odpowiedzialności sztucznego podmiotu.

Ostatni rozdział rozprawy doktorskiej zatytułowany „Funkcje poznawcze robotów humanoidalnych" rozpoczyna się od analizy pojęcia zdolności poznawczych ludzkiego podmiotu, w skład którego wchodzą percepcja, uwaga, pamięć oraz myślenie i rozumowanie. Mgr Sarowski szerzej omawia pojęcie inteligencji, w zakres której włącza (zgodnie z

klasycznymi definicjami inteligencji ogólnej) nie tylko takie dyspozycje jak umiejętność obserwacji, zdolność uczenia się, wnioskowanie, abstrahowanie, czy też rozwiązywanie problemów, ale także (zdaje się, że według własnego rozumienia) "błyskotliwość sytuacyjną", czy też „poczucie piękna i estetyki" (s. 105). Do tak szeroko zakrojonego problemu inteligencji dodaje również zagadnienia emocji i świadomości. Od charakterystyki ludzkiej inteligencji przechodzi następnie do rozpatrzenia zagadnienia kluczowego – możliwości (w tym stopnia i skutków jej realizacji) implementacji technicznej wyróżników tej inteligencji w sztuczne systemy poznawcze. Jest ono omówione na przykładzie tzw. uczenia maszynowego, w którym humanoidalne roboty zdają się dysponować takimi zdolnościami poznawczymi, jak rozpoznawanie mowy, rozumienie wypowiedzi i przypisywanie im stosownego znaczenia, czy też wytwarzanie przez maszynę wypowiedzi zrozumiałych dla człowieka. Rozpatrzona zostaje również kwestia lingwistycznych kryteriów, jakie wchodzą w rachubę w ocenie stopnia językowych zdolności poznawczych humanoidalnych robotów – wieloznaczność, nieostrość, amfibolia językowa, eliptyczność czy ekwiwokacja wypowiedzi. Rozdział kończy się rozpatrzeniem ważnego, szeroko dyskutowanego w kognitywistyce i sztucznej inteligencji, zagadnienia testu, jakiemu można byłoby poddać humanoidalnego robota, aby wykazać, iż dysponuje on zdolnościami poznawczymi takimi, jakimi rozporządza człowieka. Autor rozprawy szeroko i ze znawstwem tematu omawia historię powstania tzw. testu Turinga oraz dyskusję toczącą się wokół niego; sięga nawet do analogicznych prób sformułowanych już przez Kartezjusza i La Mettriego. Pokazuje cele takiej weryfikacji (głownie porównania co do jakości) dyspozycji ludzkich i maszynowych, jak również niekonkluzywność wyników testu Turinga w odniesieniu do poznawczych dyspozycji humanoidalnego robota, zwłaszcza jeśli rozpatrzy się jego formy ucieleśnienia, w tym potencjalne role lub płeć, jakie mógłby przyjmować (imitować). Autor sięga do literatury tego zagadnienia (P. Łupkowskiego. J. Genove czy J. Searle'a), w tym również do argumentacji St. Lema.

Podsumowując tematykę całego rozdziału, Autor stwierdza, że terminem „inteligencja" można objąć szeroki zakres przedmiotowy (ontologiczną dziedzinę jego desygnatów), w którym obok człowieka może funkcjonować także maszyna oraz zwierzę. Operacyjne znaczenie „inteligencji" proponuje zawęzić do rozwiązywania określonych problemów. Idąc za tą trafną sugestią, formułuje ważkie filozoficzne zagadnienie – czy operacyjne (formalne, czysto funkcjonalistyczne) ujęcie inteligencji właściwe dla teorii sztucznej inteligencji jest w jakimś stopniu podobne do Platońskiej i Kartezjańskiej koncepcji niematerialnej duszy, jako tej, która samodzielne funkcjonuje bez związku z ciałem? Problem ten Autor rozpatruje już tylko skrótowo (zaledwie go rejestrując), jako dwie filozoficzne kwestie: (1) ujęcia podobieństwa tych koncepcji traktujących o ogólnej dyspozycji poznawczej podmiotu „w kontekście jego bytowania w substancji fizycznej, która nie ma właściwości determinujących" (s. 116); a także (2) przyjęcia „niewystarczalności związku podmiotu poznania np. z *hardware'em* komputera bez szczególnej analizy jego roli w procesach poznawczych" (s. 117) jako rozwiązania błędnego. Uwagi te są trafne i dobrze sformułowane, nie są jednak rozwinięte ani szerzej omówione; ich filozoficzna treść nie została w pełni wykorzystana, co jak na rozprawę filozoficzną jest wadą.

W „Zakończeniu" rozprawy doktorskiej mgr Sarowski stwierdza, że jej głównym celem jest wypracowanie aparatury pojęciowej koniecznej dla scharakteryzowania społecznej natury podmiotowości humanoidalnych robotów. Dlatego też ponownie przywołuje zestaw własności, które uznaje za charakterystyczne dla niej – ucieleśnienie, komunikacyjność, autonomiczność oraz funkcje poznawcze, głównie zaś uczenie się. Nic nowego już do ich charakterystyki nie dodaje. Podkreśla niemniej (jak robił w poszczególnych rozdziałach), że o uznaniu tych własności za realizowane przez roboty decydują czynniki nie tylko obiektywne, jak zaawansowanie techniczne robotów, ale także podmiotowe, jak interpretacje, nastawienia poznawcze teoretyków czy użytkowników robotów. Zauważa nawet, że „element złudzenia

posiadania powyższych własności może działać na rzecz uznania robota za istotę żywą", co potwierdza przytoczona opinia Sherry Turkle (autorki notabene dość często w rozprawie cytowanej, chociaż nie w pełni wykorzystanej). Za niewykorzystany temat należy również uznać wspomnianą zaledwie (zresztą niezbyt zręcznie określoną) „robofilozofię", w ramach której Autor zdaje się sytuować wszystkie wątki swojej rozprawy, choć tego nie zebrał w jakieś zgrabne podsumowanie, na które w Zakończeniu byłoby miejsce.

Redakcja i język rozprawy

Rozprawa składa się z pięciu rozdziałów, z których każdy ma kilka (od dwóch do trzech) paragrafów. Rozdziały poprzedzone są Wstępem (7 i ½ stronicowym) a podsumowane Zakończeniem (zaledwie 4 i 1/ stronicowym); liczy łącznie (bez bibliografii, z wydzieloną netografią, zajmującej 10 i ½ stron) 122 stron. Układ treści w spisie jest dość czytelny, choć zdarzają się niekiedy powtórzenia (zwłaszcza o ucieleśnieniu). Rozdział pierwszy, zapowiedziany jako „terminologiczne założenia", ma o wiele szerszy charakter, gdyż zawiera także definicje kluczowych kategorii i ich klasyfikacje. Rozdział piąty ma w tytule nazbyt skrótowe określenie „funkcje poznawcze", podczas gdy mówi się w nim o zdolnościach i kompetencjach poznawczych. Bibliografia jest podzielona na dwie części: tradycyjne źródła (książki i artykuły) oraz na tzw. netografię; jest raczej obfita, jeśli chodzi o ilość, nie do końca jednak reprezentatywna, gdyż brak w niej szerszych opracowań z cybernetyki i najnowszej filozofii techniki. Spora część pozycji jest względnie stara, co nie jest być może błędem doboru źródeł, a najwyżej efektem niezbyt szerokiego oczytania w literaturze przedmiotu (bardzo zresztą interdyscyplinarnego). Koncepcje i teorie, do których Autor się odwołuje omówione są na ogół poprawnie i ze zrozumieniem, jakkolwiek nie zawsze są one pogłębione, często są pobieżne i nie wykorzystane w pełni (np. koncepcje i badania S. Turkle). Cytowane z języka angielskiego fragmenty tekstów są poprawnie przełożone. Rozprawa jest napisana językiem polskim poprawnym; jej styl jest niemniej w wielu miejscach dość zawiły i utrudnia pełne zrozumienie intencji Autora, zwłaszcza, gdy formułuje swoje rozwiązanie analizowanych hipotez.

Ocena końcowa i wniosek

Biorąc pod uwagę wartość poruszanej w rozprawie doktorskiej problematyki mówiącej o złożonym uwikłaniu relacji człowieka z robotami, jej rosnącą aktualność, w tym zwłaszcza wyraźnie sformułowaną tezę i analizę jej konsekwencji dla filozofii człowieka i techniki, a także interdyscyplinarny charakter rozprawy, można stwierdzić, że spełnia ona wymagania stawiane rozprawom doktorskim w stopniu zadowalającym.

*

Konkludując powyższe uwagi o całości pracy, uwzględniając przy tym wszystkie aktualne przepisy stanowiące o rozprawach doktorskich i przeprowadzania stosownych procedur, wnoszę o dopuszczenie mgr Łukasza Sarowskiego do dalszych kroków w procedurze nadawania stopnia doktorskiego z dyscypliny filozofia.

Rafał Hetmański