Dawid Lubiszewski

Czy sztuczne systemy poznawcze są racjonalne?

Prace Naukowe Akademii im. Jana Długosza w Częstochowie. Filozofia nr 6, 69-76

2009

PRACE NAUKOWE Akademii im. Jana Długosza w Częstochowie
Seria: Filozofia 2009, z. VI

Dawid Lubiszewski

Czy sztuczne systemy poznawcze są racjonalne?

1. Sztuczne systemy poznawcze

Systemem poznawczym nazywać będę każdą jednostkę wchodzącą w interakcję ze środowiskiem i posiadającą o nim pewną wiedzę. Tak zdefiniowany system poznawczy nazywany jest też agentem (*cognitive agent*), czyli usytuowaną, zdolną do aktywnego działania w środowisku jednostką. Cechą agenta jest to, że jego działanie zależy od otoczenia, w którym przebywa, i własnego cielesnego uposażenia[1]. Systemy poznawcze podzielić można na dwie kategorie, które zostały wyróżnione w oparciu o pochodzenie agenta. Pierwszą z nich są biologiczne systemy poznawcze, drugą zaś sztuczne systemy poznawcze (*artificial cognitive systems* lub *non-biological cognitive systems*; dalej w skrócie SSP). Badania nad zdolnościami poznawczymi biologicznych agentów, w ramach kognitywistyki, skupione są przede wszystkim na człowieku[2]. Natomiast, jak wskazuje nazwa, jako sztuczne systemy poznawcze definiuje się agentów niemających biologicznego pochodzenia, czyli stworzonych przez człowieka, bądź inny sztuczny system poznawczy. Sam zakres pojęciowy SSP jest bardzo szeroki i dlatego posługuję się podziałem na roboty i systemy sztucznej inteligencji (dalej w skrócie SSI). Podział ten nie jest całkowicie rozłączny, gdyż wraz z rozwojem technologicznym i pracami nad sztuczną inteligencją czy robotami humanoidalnymi łączy się ze sobą dwa sztuczne systemy.

---

[1] H.G. Stork, *Cognition and* (*Artificial*) *Cognitive Systems. Explanatory & exploratory notes*, dostępne online: http://www.cikon.de/Text_EN/CogNotes1.pdf, s. 1 [stan z: 1 lutego 2009 r.]

[2] Przeprowadzane są one również, choć w mniejszym stopniu, na innych organizmach żywych, poczynając od jednokomórkowych, a na małpach człekokształtnych kończąc.

Roboty, w przeciwieństwie do SSI, mają kontakt ze światem zewnętrznym (rzeczywistym). Od robotów wymaga się podejmowania działań w rzeczywistym środowisku, w którym funkcjonują, oraz reagowania na konsekwencje tych działań. Natomiast systemy sztucznej inteligencji są programami komputerowymi. Do tej grupy na przykład należą systemy eksperckie, które są programami posiadającymi specjalistyczną wiedzę z określonej dziedziny[3]. W ramach posiadanej wiedzy potrafią one dokonywać wyborów, przygotowywać propozycje rozwiązań problemów czy oceny pewnych sytuacji. Istotnym dla systemów eksperckich jest fakt, że rozwiązują one problemy, do których nie stosuje się precyzyjnych algorytmów. Dlatego efekty pracy systemu eksperckiego są podobne do tych, jakie uzyskują ludzcy eksperci z wybranych dziedzin. Współpraca pomiędzy człowiekiem a systemem eksperckim polega na dialogu za pośrednictwem klawiatury, myszki bądź mikrofonu i głośników[4]. Z tego powodu, mimo iż taki system doskonale sprawdza się, zastępując człowieka, na przykład jako informacja telefoniczna umożliwiająca rezerwację biletów czynna przez całą dobę, to jego działanie jest nieporównywalnie mniej związane z otoczeniem niż w przypadku robotów. Ponieważ SSI jest programem, który może być realizowany na różnym sprzęcie, na przykład w komputerze osobistym, to sama interakcja sprzętu z otoczeniem nie powinna wpływać na działanie programu[5]. Jak już zaznaczyłem wcześniej, podział na SSI i roboty nie jest jednak całkowicie rozłączny, gdyż nieustające prace nad sztuczną inteligencją doprowadziły do powstania hybryd, czyli sztucznych systemów, które łączą w sobie cechy systemu eksperckiego z robotem. Hybrydą jest system ekspercki zależny od sprzętu na jakim jest realizowany, na przykład system sterujący robota, czyli jego „m ó z g”[6]. W tej pracy jednak takie hybrydy traktować będę jako roboty, gdyż system ekspercki jest częścią robota.

Powyższy podział sztucznych systemów na SSI i roboty jest więc oparty o związek, jaki zachodzi pomiędzy SSP a światem, w którym działają. Dla robotów jest nim rzeczywistość realnie istniejąca, natomiast dla SSI jest nim wirtual-

---

[3] R. Forsyth, *Expert systems. Principles and case studies*, Chapman & Hall, Bristol 1989, s. IX.

[4] W. Miszalski, Z. Świątnicki, R. Wantoch-Rekowski, *Inteligentne roboty wojskowe*, Dom Wydawniczy Bellona, Warszawa 2001, s. 79.

[5] Przynajmniej w tym sensie, że zmiany w środowisku, czy wszelkie interakcje sprzętu ze środowiskiem, nie powinny wpływać na realizację programu, to znaczy, jeśli komputer zostanie przesunięty, to program nadal powinien rezerwować bilety niezależnie od tej zmiany, oraz zmiany te nie powinny wymuszać ponownego przystosowania programu do pracy. Oczywiście wszelkie przypadki losowe, jak pożary, trzęsienia ziemi czy awarie, będą wpływać na działanie programu, jeśli sprzęt na którym jest on realizowany, będzie narażony na działanie tych przypadków. Jednakże, w tej sytuacji interakcja ze środowiskiem jest bardzo ograniczona, bo tylko jednostronna. Zniszczenie sprzętu zatrzymuje działanie programu (z reguły go niszczy).

[6] Tamże, s. 79.

na rzeczywistość, czyli świat stworzony przy wykorzystaniu technologii informatycznej. Z tego powodu termin agent używany jest również do opisu programów komputerowych, gdyż istnieje on i funkcjonuje w wirtualnym świecie[7].

2. Epistemologia androidów i nauka o androidach

Epistemologia androidów (*android epistemology*) to interdyscyplinarny kierunek badawczy, który powstał w wyniku połączenia pracy, wspólnych wysiłków filozofów i inżynierów zajmujących się poznaniem[8]. Do tej grupy dołączyli też psychologowie i inni badacze zajmujący się kognitywistyką. Sam termin e p i s t e m o l o g i a a n d r o i d ó w został wprowadzony przez Clarka Glymoura, a książka pod jego redakcją, pod takim właśnie tytułem[9] wydana została w 1995 roku. Przedmiotem badań tej nowej dziedziny są możliwości i ograniczenia poznawcze sztucznych systemów. W odróżnieniu od innych dziedzin nauki badającej sztuczne systemy poznawcze, takich chociażby jak robotyka, w epistemologii androidów stawia się podobne pytania do tych, które od wieków stawiają filozofowie w teorii poznania. Jednakże w przeciwieństwie od t r a d y c y j n e j epistemologii, gdzie w centrum tych pytań był człowiek, w epistemologii androidów to miejsce zajęły SSP. Zasadniczo jednak stawiane pytania są bardzo podobne i dotyczą przekonań, wierzeń, racjonalności czy stanów mentalnych.

Za jednego z prekursorów tej dziedziny uznać można Alana Turinga, który analizował cechy, jakie powinien mieć sztuczny system, by uznać go było można za myślący. Zaproponował on test, nazywany później od jego nazwiska testem Turinga, który to miał określać zdolność maszyny do posługiwania się językiem naturalnym oraz opanowania przez nią umiejętności myślenia podobnego do myślenia ludzkiego. Test Turinga porównywał wyniki, jakie osiągnął sztuczny system rywalizujący z człowiekiem w grze w udawanie[10]. Postawione przez Turinga pytanie, c z y m a s z y n y m o g ą m y ś l e ć ? rozpoczęło dyskusję nad możliwościami poznawczymi sztucznych systemów. Tak więc,

---

[7] J.M. Bradshaw, *An introduction to software agents*, [w:] *Software agents*, MIT Press, Cambridge 1997, s. 7.

[8] M. Miłkowski, *Naturalized epistemology and artificial cognitive systems*, dostępne online: http://marcinmilkowski.pl/downloads/knew/naturalized-epistemology-artificial-cognitive-systems.pdf, s. 2 [stan z: 1 lutego 2009 r.]

[9] Książka ukazała się pod tytułem *Android Epistemology*, red. K. Ford, C. Glymour, P. Hayes, MIT Press, Cambridge 1995.

[10] A. Turing, *Maszyna licząca a inteligencja*, [w:] *Filozofia umysłu*, red. B. Chwedeńczuk, Aletheia Spacja, Warszawa 1995, s. 271–272.

jednym ze sposobów badania SSP jest porównywanie ich zdolności poznawczych z innymi systemami, w tym sztucznymi i biologicznymi. Natomiast to porównywanie może być przedmiotem badań innej nauki – nauki o androidach.

Nauką o androidach (*android science*) nazwany został interdyscyplinarny kierunek badający ludzkie poznanie i zachowanie w oparciu o sztuczne systemy poznawcze. Kierunek ten zakłada, że SSP wchodzące w interakcje z człowiekiem wywołują u niego podobne bądź takie same reakcje, jakie pojawiają się, gdy do czynienia mamy z interakcją pomiędzy samymi ludźmi. Stosowanie SSP w eksperymentach badających interakcje społeczne jest bardzo pomocne, gdyż sprawowanie kontroli nad SSP jest dużo prostsze niż sprawowanie kontroli nad człowiekiem. Dlatego nauka o androidach angażuje się w proces tworzenia i weryfikacji hipotez stawianych przez społeczne i kognitywne nauki za pomocą interakcji człowieka z SSP[11].

Różnica pomiędzy epistemologią androidów a nauką o androidach jest więc taka, że ta pierwsza bada możliwości poznawcze SSP, a druga w oparciu o SSP bada możliwości poznawcze człowieka. Jednakże owocna współpraca pomiędzy obiema dziedzinami nie jest wykluczona, co więcej, jest ona wskazana, gdyż z dokładnym poznaniem działania biologicznych istot wiąże się nadzieja na lepszą konstrukcję sztucznych istot, docelowo takich jak człowiek.

3. Racjonalność jako podstawowa cecha sztucznego systemu poznawczego

Racjonalność w sztucznych systemach poznawczych można badać na wiele sposobów. Jeden z nich, któremu został poświęcony ten tekst, traktuje racjonalność jako podstawową cechę każdego poprawnie funkcjonującego sztucznego systemu poznawczego. Inny bada związek pomiędzy racjonalnością a inteligencją. Natomiast kolejny pyta o warunki, jakie musiałby spełnić sztuczny system, by zostać uznany za moralny, i w tym kontekście rozważana jest racjonalność.

W zależności od tego, czy SSP będzie wykonywał pewne zadania, wchodząc w interakcje albo z otoczeniem rzeczywistym, albo wirtualnym, oraz jaki typ zadań będzie realizował, wyróżnić można dwa kryteria racjonalności. Z tymi kryteriami związany jest podział agentów na antropomorficznych (*anthropomorphic agent*) i zorientowanych na cel (*goal-oriented agent*). Istotną cechą tych pierwszych jest to, że ich działanie wspomaga działanie ludzi. Dlatego sposób ich funkcjonowania musi być zgodny z przyjętymi przez ludzi zasadami racjonalno-

[11] K.F. Macdorman, *Introduction to the special issue on android science*, „Connection Science", Vol. 18, No. 4, December 2006, s. 313.

ści i tak oceniany, na przykład: racjonalny agent to taki, który postępuje zgodnie z regułą odrywania. Natomiast istotną cechą agentów zorientowanych na cel jest wykonywanie określonych zadań w świecie rzeczywistym bądź wirtualnym. Od tych agentów nie wymaga się postępowania według przyjętych reguł racjonalnego zachowania, lecz wykonania zadani. Racjonalnym agentem jest taki, który prawidłowo wykonał swoje zadanie[12]. Niezależnie od tego, z jakim systemem poznawczym mamy do czynienia, będzie on spełniał jedno ze zdefiniowanych poniżej kryteriów. Dlatego można traktować racjonalność jako podstawową cechę każdego sztucznego systemu poznawczego. Poniżej przedstawiona została analiza dwu sposobów definiowania racjonalności.

Pierwsze ujęcie zakłada, że racjonalny agent postępuje zawsze w taki sposób, by w oparciu o posiadane informacje uzyskać jak najlepszy rezultat swoich działań. Jeśli więc biologiczny organizm jest głodny, to powinien on postąpić tak, by zlikwidować uczucie głodu w możliwie najkrótszym czasie. Natomiast jeśli SSP wykonuje jakąś określoną funkcję, na przykład wyłącza światło, gdy w pomieszczeniu nikt się nie porusza, to nieracjonalnym będzie wyłączanie światła, gdy w pomieszczeniu zarejestrowano ruch, gdyż – zgodnie z posiadaną wiedzą – wtedy światło powinno być włączone. Racjonalność jest więc zdolnością do wykonania najbardziej odpowiednich, czyli poprawnych, działań w oparciu o dostępne informacje. Definicja ta zakłada koherencję pomiędzy działaniami a postawami propozycjonalnymi (takimi jak wierzenia czy preferencje) systemu poznawczego[13] i oparta jest o podstawową zasadę racjonalności zaproponowaną przez Allena Newella. Zakłada ona, że agent posiadający wiedzę o działaniach i celach, jakie może w wyniku ich przeprowadzenia zrealizować, wybierze dokładnie te czynności, które do danego celu prowadzą[14]. Co więcej, proces dobierania odpowiedniego działania w oparciu o posiadane informacje powinien przebiegać zgodnie z przyjętymi przez ludzi standardami racjonalności. Na przykład: racjonalny agent będzie wyprowadzał wnioski zgodnie z zasadami logiki klasycznej. Spełnienie tego warunku powoduje, że działania jednych systemów poznawczych stają się przewidywalne dla innych, a to umożliwia efektywną interakcję pomiędzy różnymi agentami. Zdaniem Wynna Stirlinga, skonstruowanie SSP podejmującego efektywne i budzące zaufania człowieka decyzje jest możliwe tylko wtedy, gdy działanie takiego sztucznego agenta oparte jest na

---

[12] J. Pollock, *Rational Cognition in OSCAR* [w:] „Lecture Notes In Computer Science", Vol. 1757, 1999, s. 71.

[13] M.P. Wellman, *Rationality in Decision Machines*, AAAI Fall Symposium on Rational Agency, 1995; dostępne online: http://ai.eecs.umich.edu/people/wellman/decision-machine.html [stan z 1 lutego 2009 r.]

[14] Tamże.

modelu ludzkiego zachowania[15]. Stąd też problem racjonalności może być przedmiotem badań zarówno epistemologii androidów, jak i nauki o androidach. Taką definicję racjonalności spełnia dowolny diagnostyczny system ekspercki, gdyż posiada on pewną wiedzę i reguły postępowania. Jeśli realizowanie programu nie zostanie zakłócone przez jakieś przypadki losowe, będzie on zawsze postępował racjonalnie. Jednakże, nie zawsze spełnienie tego warunku jest proste. W przypadku robotów, które otwarte są na o wiele więcej wyzwań niż program komputerowy[16], posiadanie pełnej wiedzy okazuje się warunkiem niemożliwym do spełnienia, co może wpływać na prawidłowe funkcjonowanie agenta. Problem ten dotyczy zarówno agentów-robotów, jak i agentów-programów i znany jest pod nazwą p r o b l e m u r a m[17]. Problem ten przedstawić można w formie pytania: „skąd wiemy, co w danym środowisku jest ważne dla poprawnego działania agenta, a co nie?" Aby znaleźć odpowiedź na to pytanie, trzeba przeanalizować wszystkie możliwe sytuacje, w których może znaleźć się agent, i odróżnić czynniki, które ulegają zmianie, gdy wykonywane jest dowolne działanie – od tych, których działanie nie zmienia. Z pozoru ten problem może wydać się trywialny, gdyż większość z nas nie ma problemu z rozróżnieniem tego, co się zmieni, od tego, co pozostanie bez zmian. Jednakże wyposażenie SSP w taką wiedzę wydawało się jeszcze do niedawna niemożliwe, gdyż w nieskończoność można by tworzyć różne warianty każdej dowolnej sytuacji. Dla przykładu: gdy zawiązuję buty, to nie muszę sprawdzać, czy moje spodnie nie zmieniły koloru, gdyż wiem, iż takie rzeczy się nie zmieniają przy wykonywaniu tej czynności. Jednakże to, co dla nas jest oczywiste, dla sztucznego agenta już takie być nie musi. Dlatego dowolna zmiana, która nie została przewidziana przez SSP, może spowodować zakończenie jego działania[18]. Ponadto, problem ten nie pojawiłby się w filozofii, gdyby nie jego sformułowanie w ramach teorii sztucznej inteligencji[19].

---

[15] W.C. Stirling, *Games machines play*, „Minds and Machines", Volume 12, Issue 3, 2002, s. 328.

[16] Agent-program przyjmuje tylko pewne określone dane i ma dostęp do pewnej skończonej wiedzy, poza którą stawianie diagnoz nie powinno wykraczać.

[17] Problem ten został postawiony po raz pierwszy przez Johna McCarthy'ego i Patricka Hayes'a w 1969 w artykule *Some Philosophical Problems from the Standpoint of Artificial Intelligence*. Od tego czasu sam sposób definiowania problemu ulegał różnym wariacjom. W niniejszym tekście prezentowany jest on w uogólniony sposób.

[18] Z takimi konsekwencjami często spotykają się użytkownicy różnych programów komputerowych, które przestają reagować na polecenia użytkownika, gdy ten wcześniej wyda polecenia, które nie zostały przewidziane przez ów program.

[19] M. Kamermans, T. Schmits, *The History of the Frame Problem*, dostępne online: http://staff.science.uva.nl/~bredeweg/pdf/BSc/20032004/KamermansSchmits.pdf [stan z: 1 lutego 2009 r.]

Podane powyżej kryterium racjonalności nie jest odpowiednie dla agentów nakierowanych na cel. W ich przypadku istotne jest przystosowanie się do otoczenia (*fit to reality*) i prawidłowe wykonanie powierzonego zadania. W przeciwieństwie do antropomorficznych agentów, od których wymagamy przestrzegania logicznych reguł, dla agentów zorientowanych na cel nie jest to istotne dopóty, dopóki zadania są prawidłowo wykonywane[20]. Oznacza to, że nie jest ważne, czy agent posiada prawdziwe, czy fałszywe przekonania, dopóki to, co robi, przyczynia się do realizacji wyznaczonego celu. Ponadto reguła (algorytm, metoda, zasada itp.), która została wybrana i dzięki której robot wykonał określone zadanie, nie uzyskuje statusu uniwersalnej, gdyż każdy akt przystosowywania się, za każdym razem jest nowym działaniem. Wobec tego, przyjęta raz reguła za drugim razem może okazać się najgorszą z możliwych. Mając na uwadze problem ram, od agenta nie wymaga się zdolności przystosowania się do każdego możliwego środowiska, a jedynie adaptacji do aktualnego, jednego i konkretnego otoczenia, w którym ma wykonywać określone działania[21]. Ocena architektury poznawczej agenta odbywa się według podanego w tej części kryterium racjonalności, czyli zależy tylko od tego, jak dobrze wykonywane są założone cele, a nie, czy przebiegają one z przyjętymi przez ludzi zasadami prawidłowego wnioskowania. Tak zdefiniowana racjonalność jest więc cechą wszystkich tych agentów, którzy prawidłowo wykonują powierzone im zadania, niezależnie od tego, czy wykonane są one zgodnie z przyjętymi przez człowieka kryteriami racjonalności.

Podsumowanie

Przedstawione w tym tekście zagadnienia stanowią zaledwie przyczynek do dalszych badań nad racjonalnością sztucznych systemów poznawczych. Omówione dwa kryteria racjonalności nie są jedynymi, które można implementować do sztucznych agentów. Jednakże z perspektywy definiowania racjonalności jako podstawowej cechy każdego systemu poznawczego wydały mi się one najbardziej istotne. Interdyscyplinarne badania nad racjonalnością przyczyniają się do powstawania nowych problemów nie tylko w takich naukach jak teoria sztucznej inteligencji, czy robotyka, ale również w filozofii. Co więcej, porównywanie zdolności sztucznych i biologicznych agentów, jak i konstruowanie tych pierwszych, umożliwia weryfikowanie bądź falsyfikowanie hipotez stawia-

[20] J. Pollock, *Rational Cognition in OSCAR*, „Lecture Notes In Computer Science", Vol. 1757, 1999, s. 71.

[21] M.R. Forster, *How Do Simple Rules 'Fit to Reality' in a Complex World?*, „Minds and Machines", Volume 9, Issue 4 (November 1999), s. 544.

nych w naukach poznawczych, społecznych czy filozofii. W pracach nad stworzeniem szeroko rozumianej sztucznej inteligencji podkreśla się relację pomiędzy racjonalnością a inteligencją. Z filozoficznego punktu widzenia, dla dalszych badań istotne jest zwrócenie uwagi na coraz większą autonomię sztucznych agentów oraz związki pomiędzy moralnością a racjonalnością. Sukcesy na tym interdyscyplinarnym polu badawczym skłonić mogą nieprzekonanych jeszcze filozofów do szeroko rozumianego nurtu naturalistycznego.

Summary

Are Artificial Cognitive Systems Rational?

Rationality in the artificial cognitive systems is one of many subjects which modern sciences investigate. This investigation is a part of new interdisciplinary approach like android epistemology or android science. I argue that it is possible to analyze rationality at least in three ways. First, which have been described in this article treats rationality as basic feature for each cognitive system functioning correctly. Second claims rationality as condition for other cognitive features like intelligence. The third shows connection between rationality and morality. This three approaches has contributed to interesting discussion among engineers or programmers but also among philosophers, particularly philosophers of mind. These discussions are carried not only within the confines of definite domains of sciences but also within the confines of interdisciplinary domains. Cooperation between experts from different domains allows to look at known problems with completely new way, which has been human-centered for many years.

Keywords: rationality, android epistemology, artificial intelligence, artificial cognitive systems.