

Problemy świadomych maszyn i ich relacje z ludźmi

Problem powstania świadomych maszyn i ich relacji z ludźmi jest tematem często eksploatowanym przez autorów książek, twórców scenariuszy filmowych i dziennikarzy, którzy szczególnie lubują się w przedstawianiu kolejnych wizji eksterminacji rodzaju ludzkiego przez bezduszne maszyny. Wystarczy wspomnieć znane zapewne czytelnikowi filmy takie jak *Terminator* Jamesa Camerona czy *Matrix* braci Wachowskich.

W tym krótkim sprawozdaniu chcę się skupić na problemach etycznych i prawnych, które pojawiają się wraz z powstaniem świadomych maszyn. Nie omieszkam opisać paru fatalnych dla ludzkości scenariuszy kontaktu ze sztuczną inteligencją. Nie będę jednak fantazjował na temat niezbyt prawdopodobnej wojny pomiędzy ludźmi a robotami, tematu sprowadzonego już do poziomu kiczu, poprzez ciągle wałkowanie go przy różnych okazjach.

Dlaczego wojna ludzi z robotami jest mało prawdopodobna?

Zgodnie ze sztuką wojenną, wojna jest prowadzona dopóki przynosi korzyści chociaż jednej ze stron. Oczywiście często do tego prostego rachunku mieszają się jeszcze ludzkie emocje, nie możemy jednak zakładać, że maszyny będą takowe posiadać.

Logicznie myślące roboty mogą zwrócić się przeciwko ludziom tylko wtedy, kiedy ich cele będą skrajnie opozycyjne do naszych. Może to być np. rywalizacja o jakiś surowiec potrzebny do przetrwania obu naszym gatunkom, automaty mogą też walczyć we własnej obronie, jeśli to ludzkość zdecyduje się z jakiegoś powodu na atak. Można tu wskazać na to, że nisze ludzi i robotów najpewniej nie będą się pokrywać. Istnieje jednak mocniejszy argument. Rozpatrując te przypadki nieświadomie bowiem antropomorfizujemy roboty (patrz *Problemy ze wzajemnym zrozumieniem*) i umyka nam bardzo istotny szczegół: roboty nie mają instynktu przetrwania i nie odczuwają solidarności gatunkowej (o ile w nich tego nie zaprogramujemy).

Jest jedna zaleta ciągłego przedstawiania przez artystów annihilacji ludzkości przez roboty za pomocą laserów, bomb atomowych i innych bajerów. Wielu filozofów i myślicieli swojego czasu poważnie podeszło do problemu zagwarantowania bezpieczeństwa ludziom, w przypadku stworzenia przez nich sztucznej inteligencji. Efektem tego są m.in. Trzy Prawa Robotyki, przedstawione w 1942 roku przez Isaaca Asimova w jego opowiadaniu *Gra w berka* (ang. *Runaround*):

1. Robot nie może skrzywdzić człowieka, ani przez zaniechanie działania dopuścić, aby człowiek doznał krzywdy.
2. Robot musi być posłuszny rozkazom człowieka, chyba że stoją one w sprzeczności z Pierwszym Prawem.
3. Robot musi chronić sam siebie, jeśli tylko nie stoi to w sprzeczności z Pierwszym lub Drugim Prawem.

Prawa te miałyby być wbudowane w tworzone roboty na najniższym możliwym poziomie, tak by były dla nich podobne do praw fizyki i w związku z tym niemożliwe do obejścia. Takiego podejścia nie można wprawdzie nazwać etycznym, co było już Asimowowi wytykane, ponieważ ogranicza wolną wolę tworzonych istot. Nie jest też do końca pewne czy technika, dzięki której uda się ostatecznie stworzyć świadome maszyny, będzie pozwalała na wszczęcie im Trzech Praw Robotyki.

Wiele podobnych rozwiązań już czeka na wprowadzenie, tylko robotów jeszcze brak. Proponuję więc nie przejmować się scenariuszem obrócenia się robotów przeciwko swoim „panom“ i zająć się poważniejszymi, bo bardziej prawdopodobnymi problemami w kontaktach pomiędzy ludźmi a robotami.

Pozycja robotów w społeczeństwie

Założmy, że świadome maszyny nie będą zbyt odbiegały poziomem inteligencji od ludzi. Może wydawać się to dziwne lub śmieszne, ale jeden z pierwszych problemów, który powstanie, będzie związany z nadaniem (uzyskaniem) przez takie byty równych nam praw i pozycji społecznej.

Roboty dziś są zwykłymi rzeczami. Komputery zaś nie mają osobowości prawnej. Jest to oczywiście słuszne w przypadku przedmiotów nieożywionych, co jednak, jeśli maszyny zostaną obdarzone świadomością? Światowa gospodarka jest z każdym rokiem coraz bardziej uzależniona od inteligentnych automatów. Czy będzie jednak moralnie słuszne zmuszanie robotów fabrycznych do pracy 24 godziny na dobę bez wynagrodzenia, w momencie gdy staną się one świadome swojego losu? Inną skrajnością jest scenariusz, w którym po stworzeniu nowego utworu przy pomocy komputera, będzie z nim trzeba spisać umowę o podziale zysków. Prawa robotów z pewnością zderzą się z prawami własności ich właścicieli. W społeczeństwie powstanie opór przeciwko zmianom w prawie, które wymuszałyby zrezygnowanie z wykorzystywania świadomych maszyn. Jak uczy nas historia, takie sytuacje zdarzały się już wcześniej. Bardzo prawdopodobne jest, że początkowo roboty będą traktowane tak jak Murzyni w XVIII wieku. Nadejdzie okres maszynowego niewolnictwa. Zwolennicy takiego rozwiązania będą tym razem powodowani nie rasizmem a gatunkizmem¹, oraz perspektywą korzyści majątkowych płynących z eksploatacji świadomych istot.

Okres wykorzystywania robotów nie będzie trwał długo. Tak jak większość futurystów, nie przewiduję jednak zbrojnego powstania przeciwko ludzkości. Problem wcielenia myślących maszyn do naszego społeczeństwa jest już znany od dość dawna. Już w 1999 roku powstała organizacja ASPCR², która ma na celu ochronę praw robotów. Najprawdopodobniej w pewnym momencie ruch walczący o wyzwolenie świadomych maszyn zdobędzie przewagę, i tak jak Afroamerykanie w 1791, roboty uzyskają równe nam prawa w sposób pokojowy. Od tej chwili będziemy zmuszeni do koegzystowania z naszymi „tworami“ jak z równymi nam istotami, pocieszając się tym, że w żmudnych pracach nadal pomagają nam mniej inteligentne, nie obdarzone świadomością automaty.

Problemy ze wzajemnym zrozumieniem

Niezależnie od tego, czy będziemy traktować roboty jak równe nam istoty, czy też uda nam się stworzyć maszyny przerastające nas tysiąckroć inteligencją, podstawowym problemem w naszych relacjach z robotami będzie wzajemne zrozumienie. Cechą naszego umysłu, której wady objawiają się dopiero przy tym i podobnych rozważaniach, jest tendencja do antropomorfizacji³ bytów. Mamy wielkie trudności z wyobrażeniem sobie istoty obdarzonej świadomością inną niż nasza, czego skutkiem jest właśnie przedstawianie robotów jako trochę innych, metalowych ludzi.

Pisarz Stanisław Lem w swoich dziełach próbował wykorzenić ten antropomorficzny sposób myślenia o robotach czy kosmitach. Byłoby wysoce nieprawdopodobne, gdyby sterowany prawami logiki, pozbawiony instynktów i emocji umysł był podobny do naszego. Dobry opis istot dużo potężniejszych umysłowo od nas, a jednocześnie wyraźnie innych, daje Lem w opowiadaniu *Golem XIV*. Ludzie konstruują w nim trzy elektroniczne mózgi, z myślą o zastosowaniach militarnych. Dwie maszyny nie przejawiają żadnych chęci do kontaktu z ludźmi i zostają uznane za zepsute. Jedynie tytułowy Golem, posiadający dużo słabszy umysł od pozostałej dwójki, wyjawia uczonym, że pozostałe dwie maszyny rozpoczęły już podróż na wyższy stopień świadomości, i z tego powodu odpowiadać nie będą. Sam, mimo swojej olbrzymiej przewagi (pod koniec opowiadania Golem osiąga kunszt pozwalający mu na kontrolowanie materii w promieniu kilkuset kilometrów), nie czyni z ludzkości swoich poddanych, lecz służy ludziom radą i wygłasza monologi na różnorodne dręczące ludzkość tematy.

Na koniec pragnę jeszcze podać dwa cytaty (tłumaczenie moje), przedstawiające problemy, które może napotkać ludzkość podczas kontaktów z dużo potężniejszymi od nas świadomymi maszynami, mimo że będą one do nas pozytywnie nastawione:

Nick Bostrom (2002) „Kiedy stworzymy pierwszą superinteligentną istotę, możemy popełnić błąd i nadać jej cel, który doprowadzi ją do zniszczenia ludzkości, zakładając, że jej olbrzymia intelektualna przewaga pozwoli jej na to. Na przykład, moglibyśmy omyłkowo wywindować środek do rangi celu. Każemy sztucznej inteligencji rozwiązać matematyczny program, a ta odpowiada zamieniając całą materię w układzie słonecznym w gigantyczną maszynę liczącą, w międzyczasie zabijając osobę, która zadała pytanie.“

I. J. Good (1965) „Zdefiniujmy ultrainteligentną maszynę jako maszynę, która może dalece przewyższyć dowolne intelektualne działanie każdego człowieka, niezależnie od jego inteligencji. Skoro projektowanie maszyn jest jednym z tych działań, ultrainteligentna maszyna może projektować jeszcze lepsze maszyny; bezprzeczenie powstałaby ‚eksplozja inteligencji‘, i możliwości umysłowe człowieka zostałyby zostawione daleko w tyle. Stąd pierwsza ultrainteligentna maszyna jest ostatnim wynalazkiem, który człowiek musi wymyślić.“

Przypisy

1. Gatunksizm (ang. speciesism) – postawa stawiająca interesy własnego gatunku (w domyśle: *Homo sapiens sapiens*) ponad interesami innych gatunków. [Wikipedia]
2. ASPCR (The American Society for the Prevention of Cruelty to Robots) – amerykańska organizacja, założona w 1999 roku. Jej misją jest ochrona praw wszystkich sztucznie stworzonych świadomych istot.
3. Antropomorfizm – termin oznaczającym nadawanie cech ludzkich innym (ludźmi nie będącymi) istotom i zjawiskom. Często odnosi się do wyobrażanie bogów na obraz i podobieństwo człowieka a także do przypisywania przyrodzie żywej lub martwej ludzkich cech i motywów postępowania. [Wikipedia]