

Adam Radziwończyk-Syta  
Wydział Matematyki, Mechaniki i Informatyki UW

## **Mój przyjaciel robot** czyli problemy świadomych maszyn i ich relacje z ludźmi

Jednym z najpopularniejszych elementów wszelakich wizji przyszłości są świadome maszyny. Nic w tym dziwnego dają one olbrzymie pole do popisu dla filmowców i powieściopisarze i dlatego trafiły na stałe do kultury popularnej. Jak jednak powinny wyglądać stosunki ludzi z takimi robotami? Czy powinny być one traktowane jak równorzędne istoty czy jako podgatunek? W jaki sposób ułożyć stosunki między nimi a rasą ludzką?

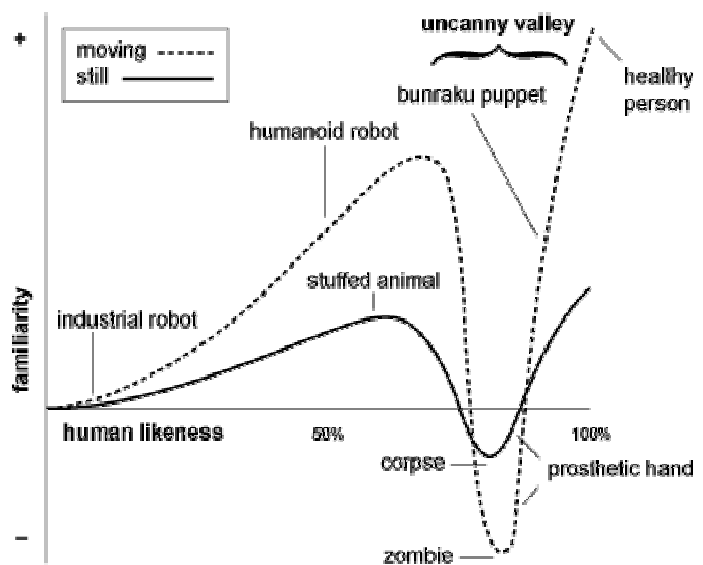
Zanim jednak zaczniemy się zastanawiać nad relacjami między ludźmi a świadomymi maszynami poświęćmy chwilę na zrozumienie czym wyróżnia się myśląca maszyna od tych spotykanych choćby dzisiaj. Podstawową różnicą jest tu znacząca różnica między zachowywaniem się przez robota tak jakby myślał, a faktyczną możliwością myślenia i samoświadomością. Jakiegokolwiek obiektywne testy, choćby test Turinga, nie pozwalają odróżnić tych przypadków zwłaszcza, że niektórzy naukowcy są zwolennikami poglądu, że nie ma między nimi różnicy. W dalszej części tekstu będzie mowa jedynie o relacjach z maszynami spełniającymi drugie z wymienionych kryteriów.

Głównym problemem relacji człowiek - świadoma maszyna jest fakt, że powszechnie oczekuje się, że maszyny będą podległe człowiekowi i on jako ich twórca będzie istotą nadrzędną. Z tego właśnie powodu często wyposaża się świadome maszyny w różnego typu ograniczenia, najbardziej znanym przykładem są tu prawa robotyki Asimowa[1]. Uważam jednak, że te postulaty wbudowania w maszyny relacji poddania wobec człowieka jest nie tylko etycznie wątpliwe, ale i niewykonalne. Z etycznego punktu widzenia nie mamy prawa powoływać świadomej do egzystencji istoty, której jedynym celem jest służenie. Zwłaszcza, że wraz z wyposażeniem maszyny w możliwość interpretowania życzeń i obdarzeniem jej potrzebną do służenia ludziom świadomością uzyskałaby ona możliwość przeprowadzenia procesów myślowych - stawiania pytań o sens egzystencji czy nieuchronność śmierci (czy trafniej terminacji?). Człowiek będą twórcą przejąłby moralną odpowiedzialność za dostarczenie odpowiedzi na takie pytania.

Wracając do podporządkowywania maszyn, ciężko sobie wyobrazić w jaki sposób zapewnić, by zachowywały się zgodnie z wytycznymi. Niezależnie, czy zostaną stworzone poprzez skomplikowany algorytm (choćby na komputerze kwantowym), czy poprzez stworzenie sztucznego mózgu nie ma łatwego sposobu zapewnienia, by nie czyniły na przykład krzywdy ludziom. Zresztą w pierwszym przypadku, by włączyć to pojęcie do algorytmu należałoby przecież dokładnie zdefiniować co to oznacza, co samo w sobie jest zadaniem trudnym. Natomiast w drugim należałoby zmodyfikować odpowiednio model mózgu lub przeprowadzić coś, co byłoby jakimś odpowiednikiem prania mózgu, ale w obydwu przypadkach nie można być pewnym w stu procentach przewidywalnych rezultatów. Wszelkie próby pokojowego podporządkowywania istot świadomych są przecież skazane na niepowodzenie.

Z powyższych rozważań wynika więc, że prawdopodobnie należy będzie traktować roboty jako mniej więcej równorzędne istoty. Czego natomiast możemy oczekiwać z ich strony? Jedną z najbardziej znanych koncepcji związanych z myślącymi maszynami jest eksplozja inteligencji, która nastąpi po osiągnięciu technologicznej osobiwości (*technological singularity*). Argumentacja jest bardzo prosta, jeśli stworzymy maszynę inteligentniejszą od człowieka to z kolei ona będzie mogła stworzyć maszynę inteligentniejszą od siebie itd. Wzrost tej inteligencji miałby odbywać się w tempie wykładniczym. Moim zdaniem, najważniejszym faktem, który zostaje pominięty przy tej analizie jest to, w jaki sposób maszyna byłaby inteligentniejsza od człowieka. Czy potrafiłaby odpowiedzieć na każde pytanie? - to nawet dziś potrafimy przy pomocy internetu. Czy rozwiązywałaby szybciej równania? - to potrafi zwykły kalkulator. Natomiast ograniczenia jest procesów myślowych wynikałyby przecież bezpośrednio z tego, że modelowałyby działanie ludzkiego mózgu. Myśląca maszyna miałaby więc cały czas takie same ograniczenia jak ludzki mózg, być może mogłaby szybciej przeprowadzać pewne procesy, ale według mnie niewiele ponadto. Robot mógłby mieć sprawniejsze "ciało" i "pamięć" niż te posiadane przez zwykłego homo sapiens, ale być może w tym momencie historii człowiek już byłby tak zintegrowany z maszynami (może poprzez implanty czy też sterowanie komputerem poprzez myśl), że to nie miałoby znaczenia. Czyli także w tym aspekcie świadome maszyny okazują się partnerami dla ludzi.

Warto się jeszcze zastanowić, jak ludzie reagowaliby na myślące maszyny w kontaktach bezpośrednich. W tego typu stosunkach najważniejszą rzeczą jasną byłby wygląd robota, a nie jego możliwości intelektualne - w szczególności stopień podobieństwa robota do człowieka determinowałby reakcję. Bardzo dobrym opisem tego zjawiska jest według mnie opis Moriego [2][3], zgodnie z którym akceptacja robota najpierw rośnie wraz ze wzrostem jego podobieństwa do człowieka, gdy przeważają cechy robota - każdy przejaw humanizmu sprawia, że robot jest bardziej akceptowany. Następnie, po przekroczeniu pewnego progu, akceptacja dramatycznie spada - gdy elementu ludzkie dominują, nieliczne cechy robota stają się bardzo zauważalne i wywołują uczucie niepokoju poprzez



dziwaczność – tak zwana „dziwna dolina” (*the uncanny valley*). Wreszcie, gdy robot staje się prawie nieodróżnialny od człowieka jest bardzo akceptowany. Ale co to oznacza dla świadomych maszyn? Według mnie, wynika z tego, że by człowiek pozytywnie na nie reagował wzory maszyn musiałyby być w okolicach szczytu przed „dziwną doliną” – musiałyby mieć kształt humanoidalny lecz elementy wyraźnie je odróżniające. W takim wypadku jednak łatwo stworzyć podział ludzie-maszyny. Zaś druga możliwość jest taka, że roboty musiałyby być nieodróżnialne od ludzi – w szczególności podobnie reagować na emocje i mieć podobną mowę ciała, w istocie musiałyby odczuwać emocje!

Podsumowując, według mnie jasne jest, że jedyny sposób, w jaki można ułożyć relacje z prawdziwie myślącymi maszynami, to partnerstwo, a one powinny starać zaadoptować się do społeczeństwa. Prawdopodobnie jednak to od stopnia, w jakim człowiek będzie zintegrowany z maszynami, zależeć będzie jak łatwo taka hipotetyczna asymilacja mogłaby przebiegać.

**Bibliografia:**

[1] [http://en.wikipedia.org/wiki/Laws\\_of\\_robotics](http://en.wikipedia.org/wiki/Laws_of_robotics)

[2] [http://en.wikipedia.org/wiki/Uncanny\\_Valley](http://en.wikipedia.org/wiki/Uncanny_Valley)

[3] <http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html>