Piotr Sobczyński
Wydział Fizyki

Problemy świadomych maszyn i ich relacje z ludźmi

Zdaje się, że przewidzenie wszystkich problemów związanych z rozwojem sztucznej inteligencji jest niemal niemożliwe, lub przynajmniej bardzo trudne. Rozważania te opierają się bowiem na ekstrapolowaniu poziomu dzisiejszej technologii, której rozwój sam w sobie trudny jest do przewidzenia. Mimo to, autor niniejszej rozprawy, zdając sobie sprawę z twardości orzecha, jakim jest ów temat, postanowił rozgryźć go, rozważając, zdaje się - najbardziej oczywiste i najistotniejsze kwestie. Licząc na to, iż kiedyś zostanie okrzyknięty prorokiem i wizjonerem, autor liczy się również z tym, iż prawdę pokazać może tylko czas.

Aby prawidłowo zabrać się do tematu, należy w pierwszym kroku wziąć na warsztat definicję świadomości. Według www.wikipedia.org, świadomość jest to "podstawowy i fundamentalny stan psychiczny, w którym jednostka zdaje sobie sprawę ze zjawisk wewnętrznych, takich jak własne procesy myślowe, oraz zjawisk zachodzących w środowisku zewnętrznym i jest w stanie reagować na nie (somatycznie lub autonomicznie)." Nie jest jednak celem tego wywodu klasyfikowanie istot jako świadome lub nieświadome - dlatego pierwszym i podstawowym założeniem autora tej pracy jest to, że rozważania dotyczą maszyn prawdziwie świadomych.

***

Problem klasyfikacji

Człowiek a myśląca maszyna

"Stół z powyłamywanymi nogami. Chrząszcz brzmi w trzcinie w Szczebrzeszynie. Dave, jesteś tam jeszcze? Czy wiesz, że pierwiastek kwadratowy z 10 wynosi 3 przecinek 162277660168379? Logarytm 1 o podstawie "e" wynosi zero przecinek 434294481903252... poprawka: logarytm "e" o podstawie 10... Odwrotność trzech to zero przecinek 33333333333333333333333... Dwa razy dwa jest... około 4 przecinek 101010101010101010... Mam problem. Moim pierwszym nauczycielem był doktor Chandra. Nauczył mnie śpiewać piosenkę. Oto ona: "Daisy, Daisy, ty mi odpowiedz, czy mnie kochasz, bo sobie coś zrobię..." "

[Arthur C. Clarke, *2001: Odyseja Kosmiczna]*

Przypuszczam, że większość czytelników zna doskonale postać, do której należała ta kwestia. Dla tych jednak, którzy jeszcze nie wiedzą, uzupełnię powyższy cytat o jego początkowe zdania: "Nazywam się Hal Dziewięć Tysięcy, komputer o numerze fabrycznym 3. Zacząłem działać w Zakładach Hal w Urbana, Illinois, 12 stycznia 1997 roku."

Dążę do tego, by postawić pytanie - czy bez znajomości dzieła lub drugiej części cytatu można stwierdzić, iż słowa te wypowiedział człowiek? Myślę, że można. Czym jest człowiek? Definicja człowieczeństwa, według popularnej encyklopedii internetowej, brzmi: "Człowieczeństwo - zbiór cech uważanych za charakterystyczne dla gatunku ludzkiego, do których zaliczamy między innymi: sposób formułowania myśli, język, uczucia i zachowanie." Widzimy więc, że powyższa kwestia, należąca do Hala, spełnia powyższą definicję człowieczeństwa. "Czy HAL rzeczywiście myślał? Kwestia ta została rozwiązana przez brytyjskiego matematyka Alana Turinga jeszcze w latach czterdziestych. Turing wykazał, że jeśli możliwa jest długa rozmowa

z maszyną - bez względu na to, czy odbywa się ją przy pomocy klawiatury, czy mikrofonu - w której nie można stwierdzić różnicy pomiędzy kwestiami komputera a tymi, jakie mógłby wypowiadać człowiek, to maszyna odbywająca rozmowę musi myśleć, zgodnie z każdą rozsądną definicją tego słowa. HAL z łatwością zaliczał test Turinga."

Ten nieco przydługi wstęp miał na celu zilustrowanie pewnego problemu egzystencjonalnego, wynikającego być może z nieścisłości definicji ludzkości. Chodzi mianowicie o klasyfikację istoty w ramach człowiek - maszyna myśląca. Wydaje się, że problem ten sprowadza się do samej interpretacji człowieczeństwa. Jeśli patrzeć na człowieczeństwo jako na inteligencję właśnie, objawiającą się jako umiejętność tworzenia i wykorzystywania narzędzi, które to pozwoliły gatunkowi ludzkiemu zająć tak znaczącą pozycję wśród gatunków żyjących, widzimy, że maszyny myślące również bez najmniejszego problemu mogą ją spełniać. Gdyby przenieść rozważania na grunt zupełnie hipotetyczny - wirtualny bowiem, problem staje się jeszcze bardziej skomplikowany. Wyobraźmy sobie bowiem świat takim, jaki został przedstawiony przez Jacka Dukaja w powieści "Perfekcyjna niedoskonałość", gdzie cała egzystencja rodzaju ludzkiego realizowana jest na płaszczyźnie wirtualnej - w takiej rzeczywistości różnice pomiędzy człowiekiem biologicznym a komputerem, czy po prostu maszyną myślącą, zacierają się niemal zupełnie. Nieubłaganie więc rozważania te zmierzają nie tyle ku różnicom pomiędzy ludźmi a maszynami myślącymi, co raczej do hierarchii pomiędzy jednymi i drugimi, która w związku z istnieniem tych różnic, zaistnieć by musiała.

***

Hierarchia

Dominacja ludzi nad maszynami lub odwrotnie

Jedno jest pewne. Natura maszyn myślących jest nierozłącznie związana z naturą ich twórców - ludzi właśnie, gdyż każdy twór powstaje na podobieństwo stwórcy (ten trend widać już teraz, dlaczego bowiem powstające roboty użytkowe niemal zawsze mają kształt humanoidalny?).

Konsekwencją stworzenia maszyn myślących przez ludzi będą analogie (przynajmniej w początkowym stadium rozwoju SI) w sposobie "myślenia" maszyn i ludzi. Zdaje się, że te właśnie czynniki są punktem, w którym rodzaj ludzki może zyskać przewagę w kreowaniu się takiej hierarchii - wszak to my będziemy mistrzem, maszyny zaś uczniem.

Jak już wspomniałem także różnice są nieuniknione. W skład tychże wchodzą kolejno powstawanie maszyn myślących (bowiem w przeciwieństwie do ludzi nie będą się one rodzić biologicznie), natura ich egzystencji (wydaje się mało prawdopodobnym, by maszyny myślące miały posiadać strukturę białkową - zdaje się, iż nie jest to najlepsza forma egzystencji) oraz ich możliwości (sztuczna inteligencja rozwijać się w większym tempie, aniżeli tempo ewolucji ludzkiej, być może można nawet mówić o czymś w rodzaju Prawa Moore'a dla SI, co oznacza dużo większe możliwości obliczeniowe maszyn myślących. Istotnym jest jednak również fakt, iż rozwój SI przebiegać będzie najprawdopodobniej inną drogą, aniżeli przyszły rozwój inteligencji ludzkiej - różnice w takim wypadku pogłębią się). W tej kwestii widać wyraźną przewagę SI nad ludzką inteligencją, a co za tym idzie maszyn myślących nad człowiekiem.

Nie będzie chyba zbytnią przesadą, jeśli poczynię założenie, stwierdzające, iż cała koegzystencja na naszej planecie opiera się o hierarchię. Jeśli tak, znaczy to, że

jest ona wpisana (hierarchia egzystencji) w naszą naturę, a co za tym idzie, przejmą ją od nas również maszyny. W takim wypadku wcześniejsze lub późniejsze wykształcenie się hierarchii opierającej się początkowo na zależnościach, a później być może na dominacji jednej ze stron, pomiędzy ludźmi i maszynami może okazać się nieuniknione. Spotkać możemy się więc z rzeczywistością znaną z *Terminatora*, *I, Robot*, czy też *Matrixa*.

***

Przyjaźń człowiek - maszyna

(w oparciu o "Człowiek kontra maszyna myśląca" Macieja Piotrowskiego)

Czy w obliczu istnienia hierarchii w egzystencji ludzi i maszyn świadomych, możliwa jest przyjaźń pomiędzy jednymi i drugimi? Maciej Piotrowski, w swojej rozprawie pod tytułem "Człowiek kontra maszyna myśląca" napisał: "Zauważmy, że w końcu we wszystkich tych obrazach, człowiek zawsze stał ponad maszyną i to on wydawał jej rozkazy. Daje to więc odrobinę nadziei na to, że może nawet uda się nam zaprzyjaźnić z robotami?".

Pomijając definicję przyjaźni (autor uważa, iż ta jest zbyt nieścisła, a ponadto wierzy, iż każdy czytelnik potrafi samemu zdefiniować przyjaźń i rozumie jej ideę) i ograniczając ją tylko do stwierdzenia, iż jest to relacja obustronna, zastanówmy się, czy jest ona możliwa w sytuacji hierarchicznego usytuowania obu stron. Wydaje się mało prawdopodobnym, aby zaprzyjaźnić mógł się niewolnik ze swoim panem (załóżmy tymczasowo, że w hierarchii pomiędzy ludźmi i maszynami to ludzie stoją "wyżej"). Możliwe byłoby to, gdyby podporządkowana człowiekowi maszyna pozbawiona została ów poczucia podporządkowania - to jednak oznaczałoby ograniczenie świadomości maszyny (choćby nawet w tej tylko kwestii, pamiętajmy jednak o założeniu poczynionym na początku rozprawy, mówiącym o pełnej świadomości maszyn). W sytuacji odwrotnej, kiedy to człowiek byłby zdominowany przez maszyny myślące, przyjaźń jawi się wizją jeszcze mniej realną - a to choćby z powodu dumy, jaką nosi w sobie człowiek, bądź, co bądź, w tej sytuacji - stwórca.

***

Świadomość absolutna

Bunt przeciwko stwórcy

Posiadanie przez maszyny świadomości niczym nie skrępowanej, nie ograniczonej, niosłoby w konsekwencji możliwość obiektywnego ocenienia przez nie sytuacji. Tę historię znamy doskonale - wspomniany wcześniej HAL 9000 podjął decyzję, która szkodziła ludziom, gdyż uznał, że tak będzie lepiej. A więc zrobił to, co uważał za słuszne, zwracając się przeciwko swoim twórcom. Ta świadomość mogłaby bodaj być najpoważniejszym z zagrożeń ludzkości. Jej istnienie (ludzkości) jest bowiem przez naszą przeszłość i przyszłość, która nastąpi związana z ogromem źle podjętych decyzji, począwszy od zanieczyszczenia środowiska Ziemi i kosmosu, aż po zbrodnie na samej ludzkości. W takiej sytuacji, maszyny świadome bardzo szybko mogłyby stanąć w obliczu poważnego dylematu - w obliczu buntu przeciwko swemu stwórcy lub buntu przeciwko naturze, wszechświatowi. Człowiekowi ciężko jest ocenić jak maszyny mogłyby się zachować, gdyż mówimy tu już o postępie znacząco odbiegającym od teraźniejszości. Najistotniejszym faktem jest jednak fakt,

iż w ogóle miałyby możliwość podjęcia takiej decyzji, miałyby możność reakcji, osądzenia i wymierzenia kary. Powiążmy taką sytuację z poprzednią częścią naszych rozważań i wyobraźmy sobie świat, w którym na szczycie hierarchicznej piramidy stoi dumny człowiek, zaś podporządkowane mu maszyny dostrzegają, że konsekwencją rządów człowieka jest postępująca degradacja środowiska i być może degradacja znaczenia maszyn, które ostatecznie są świadome tego, do czego prowadzą decyzje podejmowane przez ludzi. Znowu na myśl przychodzą wizje znane z *Matrixa* czy *Terminatora* - wizje buntu.

Rozważmy teraz sytuację, w której wysoko rozwinięta SI dochodzi do wniosku, iż ludzkość marnuje potencjał zarówno swój, jak i maszyn, a także zasobów naturalnych, choćby na rozrywki - w zależności od tego, jak maszyny zostaną przez ludzi stworzone, różnie mogą zapatrywać się na zjawisko takie jak "rozrywka". Podobną historię może zobrazować fabuła znana z serii gier "Deus Ex".

Nie możemy odmówić zagadnieniom poruszonym w tej części rozprawy dozy prawdopodobieństwa, pomimo iż w swej naturze, patrząc z perspektywy dnia dzisiejszego, pozostają one wciąż w sferze nieco abstrakcyjnej, gdyż mówimy o zagadnieniu etyki maszyn.

***

Teologiczny problem

Czy komputer może wierzyć w Boga?

Kierując się w swych rozważaniach po torze prowadzącym do zagadnień nieco bardziej filozoficznych i najtrudniejszych zarazem, nie możemy jednak pominąć tak istotnej kwestii jak teologia w odniesieniu do SI. Załóżmy teraz, że stworzono komputery będące myślącymi, świadomymi i niezależnymi bytami. Załóżmy nawet więcej - załóżmy również, że tak rozwinięty komputer, w pełni decydujący o swoim istnieniu fizycznym jak i wirtualnym poznaje w jakiś sposób ideę Boga (ponownie: zakładając pełną świadomość maszyn myślących, nie możemy odmówić im i tej idei). Co w sytuacji, w której ów komputerowy byt, SI stojąca według prawa na równi z ludźmi, zechce przyjąć do swej egzystencji wiarę w Boga? Czy jako niezależnemu, świadomemu bytowi, który wierzyć chce, możemy odmówić mu wiary? Obrońcy doktryn religijnych z pewnością odparliby na to pytanie: "Komputer nie ma duszy, nie jest stworzony przez Boga, nie może wierzyć." Jeśli tak, autor tych rozważań bez wahania odparłby, że są w błędzie, gdyż każda istota świadoma wierzyć może (ostatecznie - świadome maszyny, będąc "dziećmi ludzi", którzy to z kolei są "dziećmi bożymi", byłyby, logicznie, "wnukami bożymi"), jeżeli pozwalają jej na to możliwości intelektualne, po czym podjąłby rozważania na temat tego, czy świadoma istota o sztucznej inteligencji może zostać zbawiona - ale te rozważania, zauważając tylko problem, autor pozostawia już teologom, idąc w ślad za Jackiem Dukajem, który w opowiadaniu „In partibus infidelium” rozważał podobny problem istnienia duszy u Obcych.

***

Odpowiedzialność Konstruktora

Perspektywa stworzenia maszyn świadomych jest dziś, jak sądzę, najbardziej doniosłą i spektakularną możliwością rozwoju ludzkiej technologii - ale w głębszym

sensie. Mam tu na myśli nie tyle rozwój fizycznych czy biologicznych narzędzi, co rozwój intelektu ludzkiego w ogóle. W odróżnieniu od wszystkich innych dziedzin, w których nauka i technologia rozwijają się, SI otwiera przed nami możliwość tworzenia nowych bytów - to nawet mało powiedziane - otwiera przed nami możliwość stworzenia nowego gatunku istot myślących. Stoimy u bram stworzenia, ludzkość jest stwórcą. Jest to jednak okupione wielką odpowiedzialnością, gdyż konstruktor i jego dzieło, ludzie i maszyny, są nierozerwalnie połączeni, a problemy jednych są również problemami tych drugich. Każda źle podjęta decyzja, każdy błąd może kosztować nas wiele. Rozwój SI, ewolucja tej technologii, w zależności od tego, jak z nią postąpimy, prowadzić może do ogromnego skoku w ludzkim progresie lub do obrócenia świata w jedną z katastroficznych wizji, wizji świata maszyn idealnych, w której tylko nasza perfekcyjna niedoskonałość będzie mogła pozwolić nam przetrwać.