



Lapidarium



## **Recenzja książki *Moral Machines. Teaching Robots Right from Wrong***

---

Autorzy: Wendell Wallach, Colin Allen  
Wydawnictwo: Oxford University Press  
Data wydania: 2008  
Ilość stron: 288

Dawid Lubiszewski

*Etyka i roboty* stanowią parę słów, która wielu osobom wciąż kojarzy się z filmami czy opowiadaniem z gatunku fantastyki naukowej. Jednakże gwałtowny rozwój robotyki w ciągu ostatniego ćwierćwiecza spowodował, że zagadnienia wcześniej poruszane przez powieściopisarzy czy reżyserów stały się przedmiotem naukowej debaty. Wynikami toczonych dyskusji są liczne konferencje i artykuły oraz badania. „*Moral Machines. Teaching Robots Right from Wrong*” jest jedną z pierwszych książkowych publikacji traktujących o etycznych problemach związanych z rozwojem robotyki. Napisała ją ona przez dwóch amerykańskich filozofów: profesora Colina Allena z Departamentu Historii i Filozofii Nauki na Uniwersytecie Indiana oraz Wendella Wallacha z Interdyscyplinarnego Centrum Bioetycznego na Uniwersytecie Yale. Książka ta zwraca uwagę na rosnącą odpowiedzialność, jaką powierza się sztucznym agentom, zarówno wirtualnym, jak i rzeczywistym. W związku z tym konieczne jest opracowanie coraz bardziej wyszukanych rozwiązań umożliwiających podejmowanie przez sztuczne jednostki moralnych decyzji. Autorzy nie boją się postawić w książce wielu ważnych pytań i odpowiedzi. Pytają, czy etyka maszyn, nazywana też roboetyką, jest rzeczywiście potrzebna. Czy roboty mogą być moralne i co należałoby uczynić, by można je było za takie uznać? Podobnie jak inni badacze, zajmujący się roboetyką, zwracają oni uwagę, iż mimo niepowodzenia projektu Sztucznej Inteligencji, a więc niezyskania sztucznych jednostek o inteligencji zbliżonej do człowieka bądź takiej samej, pojawiające się problemy z zakresu etyki we współczesnej robotyce wymaga-

ją nowego spojrzenia. Dotychczasowe systemy etyczne czy rozwiązania, jakie wypracowane zostały przez etyków, nie dotyczyły bowiem interakcji pomiędzy ludźmi a robotami, pomiędzy robotami a innymi żywymi organizmami, jak i robotami pomiędzy sobą. Praca Allena i Wallacha jest jedną z pierwszych, w której na poważnie podejmowany jest problem moralności sztucznych istot. Składa się ona z 12 rozdziałów, w których po kolei analizowane są różne zagadnienia. Całość liczy 288 stron i napisana została przystępnym językiem, zarówno dla etyków, inżynierów, jak i osób niezwiązanych na co dzień z omawianą tematyką. Jeśli więc ktoś chciałby dowiedzieć się, z jakimi problemami spotykają się roboetycy i w jaką stronę powinny iść przyszłe prace, to z pewnością powinien zajrzeć do tej książki. Jednakże, jak wcześniej zaznaczyłem, osoby związane z tematyką nie będą rozczarowane zawartością tej pozycji. Umożliwiają to przynajmniej trzy fakty. Pierwszym z nich jest wiedza i doświadczenie, jakie posiadają autorzy w badaniach interdyscyplinarnych. Drugim mogą być pozytywne recenzje, jakie otrzymała ich książka. Trzecim – najważniejszym – jest jej zawartość. Co więc znajduje się na tych 288 stronach?

Przede wszystkim autorzy przedstawiają koncepcję autonomicznie moralnych agentów, czyli jednostek, które zdają sobie sprawę z moralnych konsekwencji swoich działań. Oczywiście przedmiotem ich badań nie jest człowiek, który był zawsze w centrum dyskusji etycznych, a robot. Rozszerzenie definicji podmiotu moralnego na sztuczne jednostki nie jest jednak zabiegiem czysto teoretycznym. Jest wręcz przeciwnie, to praktyka życia codziennego wymusiła powstanie nowej etycznej dyscypliny. Rosnąca obecność i autonomia działań sztucznych jednostek powoduje, że znajdują się one coraz częściej w sytuacji wymagającej oceny moralnej. Brak możliwości wykrycia takiej sytuacji i odpowiedniej wobec niej odpowiedzi był – i niejednokrotnie jest – przyczyną stwarzania zagrożenia życia człowieka. Zdaniem autorów nie wystarczy bowiem konstrukcyjnie ograniczyć zdolności robota do czynienia zła. Innymi słowy nie chodzi o to, by robot nie zadał człowiekowi fizycznych obrażeń przy przypadkowym kontakcie z nim, ponieważ jego wystające czy ostre części pokryte zostały jakimś materiałem. Choć podobnego typu zabiegi mogą być przedmiotem prac inżynierskich, to ich głównym celem powinno być zaimplementowanie systemu, wykrywającego na przykład sytuacje zagrożenia życia dla człowieka, będące wynikiem działań robota. Jednak zdolność do podejmowania działań bądź ich zaniechania na skutek oceny moralnej jest o wiele bardziej problematyczna w robotyce niż mogłoby się wydawać. Pokazują to autorzy na przykładzie robotów wojskowych, których dynamiczny rozwój i udział w działaniach wojennych obserwujemy w ciągu ostatnich lat. Dla robotów wojskowych celem priorytetowym jest prawidłowe wykonanie działania a nie ocena moralna swoich czynów. Ponadto autorzy odwołując się do współcześnie prowadzonych badań pokazują jak niezmiernie złożonym jest proces podejmowania moralnych decyzji u człowieka i przed jakimi wyzwaniem staje współczesna robotyka i badania nad sztuczną inteligencją. Nie wystarczy zaimplementować w postaci algorytmu jakiś konkretnych etycznych norm, sztuczna jednostka po-



winna jeszcze umieć na przykład postawić się w sytuacji drugiej osoby. Utrudnia to znacznie stworzenie sztucznych istot posiadających cechy podmiotów moralnych.

Natomiast jednym z wyzwań, przed jakimi staną filozofowie, jest odejście od ogólnych norm postępowania na rzecz konkretnych wskazówek w pewnej sytuacji. Bo wiem współczesna technologia nie pozwala na tworzenie jednostek, które zrozumiałyby ogólną zasady typu „czyń dobro”, „szanuj bliźniego”. Dlatego trzeba tworzyć bardzo konkretne normy postępowania, których zastosowanie ograniczać się będzie tylko do robotów działających w określonej kulturze i w określonym miejscu. Kolejnym poruszonym przez autorów ważnym zagadnieniem są różnice pomiędzy robotami a ludźmi, które powodują, że opracowywana dla robotów etyka może znacznie różnić się od tej, która opracowywana była przez filozofów przez ponad dwa tysiące lat. Sztuczne jednostki obecnie nie są w stanie tak szybko analizować docierających do nich informacji z otoczenia, jak czynią to ludzie. Tak jak napisałem na wstępie: książka ta nie jest kolejną pozycją z fantastyki naukowej. Tym samym znajdziemy w niej przykłady odwołujące się do współcześnie wykorzystywanej technologii. Jest to kolejną zaletą tej książki. Opisany zostaje między innymi pojawiający się w wielu dyskusjach z etyki praktycznej przykład z jadącym tramwajem. Etykom zapewne są znane różne wariacje tego przykładu, jednak ostatecznie dotyczą one tego samego. Mianowicie, którą z dwóch możliwości wybierzemy: czy pozwolimy tramwajowi przejechać torem pierwszym i przejechać jedną osobę, czy torem drugim i przejechać osób pięć. Współcześnie jednak przed takim problemem stanąć może program komputerowy zarządzający kolejką. Następuje więc drastyczna zmiana, przedtem bowiem tę sytuację mogliśmy rozpatrywać indywidualnie, jeśli rzeczywiście miała miejsce, i wysłuchać racji motorniczego, dlaczego zdecydował się na taki a nie inny ruch. W obecnej sytuacji powinniśmy jednak zabezpieczyć się na tyle, by program komputerowy wcześniej był przygotowany na taką ewentualność. Inna z podjętych kwestii to rola świadomości w moralności. Autorzy stoją na stanowisku, że warunek posiadania świadomości przez sztuczne jednostki jest zbyt rygorystyczny. Roboty już istnieją i działają w naszej sferze moralnej pomimo tego, iż świadomości nie posiadają.

Oprócz zagadnień filozoficznych, w książce podejmowane są też zagadnienia inżynierskie, dotyczące metody implementacji etycznego systemu w sztuczną jednostkę. Autorzy opisują trzy metody: (a) oddolną, gdzie programuje się gotowe normy, (b) odgórną, gdzie jednostka uczy się norm, i (c) hybrydową, która jest połączeniem dwóch wcześniejszych. Ostatniej z metod poświęcają oni najwięcej miejsca i z nią wiążą największe nadzieje. Tym samym zarówno naukowiec, jak i filozof znajdą w tej książce dla siebie coś interesującego.