

Kodeksy etyczne robotów: zagadnienie kontroli sprawowanej przez człowieka

Riccardo Campa

Instytut Socjologii Uniwersytetu Jagiellońskiego w Krakowie

Streszczenie: W niniejszym artykule pokrótce przedstawiono dwa najbardziej znane ujęcia roboetyki: Asimova „Prawa roboetyki” oraz EURON-u „Mapę drogową roboetyki”. Następnie porównano ich założenia z punktu widzenia pewnych praktycznych problemów związanych z robotyką, zwłaszcza z użyciem robotów w działaniach wojennych czy zadaniach związanych z nadzorem. Pokazano również, że w odróżnieniu od podejścia Asimova, który zasadniczo nie dopuszcza możliwości stosowania przemocy przez roboty, pięć zasad roboetyki sformułowanych przez EURON (*Safety*: robot musi być sterowany przez człowieka; *Security*: zapobiegaj niewłaściwemu lub nielegalnemu użyciu robota; *Privacy*: chroni dane przechowywane przez robota; *Traceability*: monitoruj działania robota; *Identifiability*: nadaj unikalny identyfikator każdemu robotowi) stara się raczej uregulować tę kwestię. Wreszcie wykazano, że zasady bezpieczeństwa muszą być lepiej określone i nie mogą być postrzegane jako *panaceum*. W tym kontekście zawsze należy brać pod uwagę dwa aspekty: a) autonomia robotów wzrasta wraz z postępem ich zaawansowania technicznego, b) „człowieczeństwo” jest pojęciem mglistym, należy zatem wyraźnie określić, jaka grupa ludzi powinna sprawować kontrolę nad robotami, i w jakich okolicznościach.

Słowa kluczowe: robot, humanoid, roboetyka

1. Jaka etyka robotów?

Przez wiele lat robotyka znajdowała swoje zastosowanie głównie w przemyśle motoryzacyjnym, na przykład przy montażu czy lakierowaniu samochodów. Wraz z rozwojem roboetyki wojskowej i humanoidalnej pojawiły się nieznane wcześniej problemy natury etycznej. Coraz bardziej skomplikowane maszyny wchodzą w interakcje z człowiekiem już nie tylko na ograniczonej przestrzeni linii montażowej, ale również na polach bitew bądź w domowym zaciszu, czego skutki są często nieprzewidywalne. Roboty nie tylko opuszczają mury fabryk, w których do tej pory były zamknięte, ale są coraz bardziej wyspecjalizowane i zdolne do samodzielnych działań. Dlatego właśnie społeczność naukowa dostrzegła potrzebę zinstytucjonalizowania dyscypliny, której celem jest zmierzenie się z problematyką etyki

robotów.³⁾ Jednym z podstawowych założeń roboetyki jest ustalenie sposobu i stopnia optymalnej kontroli robotów przez człowieka.

Po rozpoznaniu konieczności podjęcia takich działań, pozostaje jeszcze ustalenie, jaki kodeks etyczny powinien być wzorcem przy podejmowaniu decyzji, podpisywaniu konwencji, formułowaniu zaleceń. Byłoby przecież naiwnym sądzić, że jest zrozumiałe samo przez się, czy dana rzecz jest etyczna, czy nie. Gdyby tak było naprawdę, debata byłaby niepotrzebna. Konstruktorzy i użytkownicy robotów wiedzieliby, co robić, działaliby przecież w myśl i na rzecz dobra. Albo odwrotnie, będąc świadomymi swojego nieetycznego postępowania, działaliby w wielkiej tajemnicy. Ale tak nie jest.

Czytając artykuły dziennikarzy poświęcone nowym technologiom, takie naiwne myślenie można zaobserwować niestety dość często. Nierzadko zdarza się, że trafiamy na opiniotwórcę lub polityka, który stwierdza: „to nie jest etyczne” – bez wyjaśnienia podstaw etycznych czy przedstawienia konkretnych powodów. Wystarczyłaby jednak prosta uwaga, aby zrozumieć, iż każda ocena natury etycznej wymaga racjonalnej argumentacji: jeśli liczna grupa osób działa w pewien ustalony sposób (na przykład konstruuje roboty do celów wojennych) i robi to otwarcie, jasne jest, że przynajmniej ta grupa uznaje swoje postępowanie za akceptowalne pod względem etycznym. Mówiąc krótko, ma swoje powody. Nie muszą one być oczywiście akceptowane, ale każda krytyka musi wynikać z jasnych i racjonalnych kryteriów.

Nie możemy jednak w tym miejscu rozpocząć dyskusji na temat podstawowych teoretycznych problemów etycznych, ani też przedstawić naszej kompleksowej wizji moralności. Nie jest to celem ani głównym tematem niniejszego artykułu. Możemy powiedzieć jedynie, iż odwołujemy się do „klasycznego” pojmowania etyki, czyli, innymi słowy, zgodnego z tradycją filozofii greckiej, nieograniczającej zachowań etycznych wyłącznie do altruizmu. Oczywiście bezinteresowna postawa filantropa jest uprzywilejowaną formą etyki, ale są nią również zachowania roztropne czy zmierzające do czynienia dobra dla siebie lub dla własnej społeczności.

Mówiąc bardzo ogólnie, etyka jest znajomością i praktykowaniem „dobra” (które Arystoteles sprowadzał do

¹⁾ Aspekty techniczne por. G. Legnani, *Robotica industriale [Robotyka przemysłowa]*, CEA, Mediolan 2003, lub B. Basilio, *Modellistica dei robot industriali [Modele robotów przemysłowych]*, CELID, Turyn 2002. Zagadnienia związane z aspektami ekonomicznymi i socjologicznymi automatyki przemysłowej por. R. Campa, *Considerazioni sulla terza rivoluzione industriale [Rozważania o trzeciej rewolucji przemysłowej]* [w:], *Il pensiero economico moderno [Nowoczesna myśl ekonomiczna]*, rok XXVII lipiec-wrzesień nr. 3, Piza 2007.

²⁾ Clapper J.R., Young J.J., Cartwright J.E., Grimes J.G., Payton S.C., Stackley S.J., Popps D. (eds.), *Unmanned Systems Integrated Roadmap 2009–2034*, www.fas.org.

³⁾ Verruggio G., *La nascita della roboetica [Narodziny etyki robotów]*, w *Leadership medica*, nr 10, 2007. www.leadershipmedica.com/sommari/2007/numero_10/Veruggio/Verruggio.pdf

pojęcia „szczęście”)⁴). Tak więc u podstaw działania etycznego leży, jako główna wartość, życzliwość (chęć czynienia dobra), ale takie życzliwe działanie może być skierowane do wielu podmiotów: nas samych, własnej rodziny, własnej społeczności (miasta, partii, klasy społecznej, narodu, imperium itd.), całej ludzkości, zespołu istot czujących (ludzi, zwierząt wyższych, hipotetycznych przybyszów z kosmosu, bardzo zaawansowanych technologicznie maszyn itd.), zbioru istot żywych (obejmującego roślinność, insekty, bezkręgowce itd.), lub wręcz do całego istnienia (w tym do gór, jezior, rzek, planet, gwiazd itp.). Jeżeli weźmiemy pod uwagę formy życzliwości także wobec bytów nieludzkich, możemy wtedy mówić o etyce transludzkiej. Do dylematów etycznych dochodzi wówczas, gdy powstaje konflikt, to znaczy, gdy działanie dobroczynne skierowane do kogoś lub czegoś staje się *ipso facto* działaniem szkodzącym komuś innemu. Niestety, biorąc pod uwagę złożoność świata, sytuacje te są częstsze niż moglibyśmy sobie tego życzyć. Jedynie nieprzejednani idealisci nie zdają sobie z tego sprawy. Nieporozumienia na tle etycznym wynikają zazwyczaj z ambiwalentności wielu naszych zachowań właśnie na płaszczyźnie czynienia dobra i zła⁵).

Pouczające byłoby szczegółowe zglebienie wszystkich wysuniętych supozycji, wszystkich odnośnych stanowisk, statystycznej absolutnej wielkości każdej przyjętej postawy. Niestety ze względu na ograniczenie wynikające z miejsca, jakie mamy do dyspozycji, zmuszeni jesteśmy ograniczyć się do analizy jakościowej jedynie najczęściej spotykanych postaw w mediach i literaturze lub tych, które subiektywnie uznajemy za najbardziej interesujące.

2. Trzy prawa robotyki Asimova

Najbardziej znana propozycja dotycząca etyki robotów – i jednocześnie jedna z pierwszych – została opracowana przez Isaaca Asimova jako Trzy Prawa robotyki:

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 Prawem Pierwszym lub Drugim.

Trzy prawa Asimova po raz pierwszy ukazały się w opowiadaniu „Zabawa w berka”, później włączonym do antologii *Ja, robot*,⁶ na podstawie którego w 2004 r. powstał film o tym samym tytule, w reżyserii Alexa Proyasa, z udziałem Willa Smitha. W wyobrażeniu rosyjskiego pisarza Trzy Prawa są celowo zakodowane w pozytronowych mózgach robotów. Asimov klasę robotów postępujących

według Trzech Praw nazywa „Asenion robots”. Klasyfikacja ta zakłada istnienie robotów będących w stanie przyjąć postawy niezgodne z tymże kodeksem.

Ponadto Asimov jest mistrzem w przedstawianiu wszelkich paradoksów, nieporozumień, ambiwalencji, mogących wyniknąć w trakcie wspólnej egzystencji robotów i ludzi tam, gdzie te prawa działają. W jednym z opowiadań na przykład opisane zostały niepożądane konsekwencje Pierwszego Prawa: robot nie może pracować jako chirurg, ponieważ mógłby skrzywdzić człowieka. Jednocześnie nie może opracowywać strategii gry w footballu amerykańskim, ponieważ mogłyby być one przyczyną nieszczęśliwych wypadków wśród graczy. Aby uniknąć tych i podobnych problemów, przestrzeganie tych praw nie jest uważane za nieodzowne u robotów bardzo wysokiej klasy.

Giuseppe Bonaccorso zauważa, że „dokładna realizacja trzech praw jest bardzo często sprzeczna z zasadami moralnymi człowieka: wyobraźcie sobie, że robot bierze udział w kłótni między dwiema osobami i w pewnym momencie jedna z osób wyciąga pistolet i grozi drugiej śmiercią. Co powinien zrobić robot? Zdałoby się, że powinien interweniować w celu uratowania życia osobie bez broni, lecz nie gwarantowałyby to z pewnością powodzenia takiej akcji: obaj mogliby stać się ofiarami napastnika, który, czując się zagrożony, byłby zmuszony do strzelania, nie zwracając uwagi na konsekwencje. Dobry negocjator zareagowałby na pewno zupełnie inaczej... Żaden program nie jest w stanie ocenić w czasie rzeczywistym wszystkich możliwych hipotez i jedynie uczucie empatii (jako zdolne do wykluczenia *a priori* wszelkich opcji zbyt nieadekwatnych do sytuacji) jest w stanie przekazać ewentualnemu odbiorcy, czy to ludzkiemu czy sztucznej inteligencji, że dobre słowo jest w zupełności wystarczające, by rozbroić człowieka z bronią w rękę”⁷).

Przykład Bonaccorso jest wyjątkowo ważny dla istoty tego opracowania. Weźmy jednak pod uwagę również inną sytuację, gdy napastnik nie może zostać powstrzymany inaczej, jak poprzez jego zabicie. I właśnie z powodu Pierwszego Prawa Asimova robot może wybrać jedynie drogę negocjacji, a więc drogę błędną. A konkretnie, na przykład w odniesieniu do robotyki wojskowej, na polu bitwy robot ma za zadanie obronę swoich ludzkich kompanów przed zagrożeniem ze strony innych ludzi, a negocjacje nie wchodzi w grę. Jeśli nie może wyrządzić krzywdy człowiekowi (nawet jeśli jest on wrogiem), to jak ma walczyć? Trzy prawa Asimova zdają się być pomyślane bardziej dla świata, w którym panuje pokój, niż świata z konfliktem zbrojnym. Z definicji, niejako *a priori* zabraniają stosowania robotów do celów wojennych.

⁴) Arystoteles, *Etyka nikomachejska*, tłum. Daniela Gromska, Wydawnictwo Naukowe PWN, Warszawa 2007.

⁵) W kwestii ambiwalencji norm deontologicznych por. R. K. Merton, *Sociological Ambivalence and Other Essays*, Free Press, New York 1974. W Polsce ekspertem w tej dziedzinie jest bez wątpienia Lech Witkowski. Por. L. Witkowski, *Idea ambiwalencji*. [w:] Tenże. *Dwoistość w pedagogice* Bogdana Suchodolskiego (z aneksem o Sergiuszu Hesse), WIT-GRAF, Kraków 2001: 123–134; L. Witkowski, *Ambiwalencja w kulturze jako wyzwanie dla pedagogiki ogólnej*. [w:] Tenże: *Między pedagogiką, filozofią i kulturą. Studia, eseje, szkice*. Tom III *Tryptyku edukacyjnego*, IBE, Warszawa 2007: 90–107; L. Witkowski, *Ambiwalencja jako kategoria dla socjologii edukacji*. [w:] Tenże: *Edukacja wobec sporów o (po)nowoczesność*. Tom I *Tryptyku edukacyjnego*, IBE, Warszawa 2007: 205–215.

⁶) I. Asimov, *Runaround*, [w:] *I Robot*, Bantam Books, New York 2004 (1950): 25.

⁷) G. Bonaccorso, *Roboetica: tra fantascienza e realtà. Analisi critica delle tre leggi di Asimov* [Etyka robotów: między science-fiction a rzeczywistością. Krytyczna analiza trzech praw Asimova], 2004, <http://www.scuoladirobotica.it/roboetica.htm>.

Przyjrzyjmy się jeszcze jednemu bardziej ekstremalnemu przypadkowi: jeśli grupa terrorystów (jak Aum Shinrikiō) za cel postawiłaby sobie eksterminację całej rasy ludzkiej, co powinny zrobić roboty? Czy również w tym wypadku nie powinny nic robić, aby nie wyrządzić krzywdy ludziom (terrorystom)? Mamy tu do czynienia z sytuacją typowej ambiwalencji etycznej, o której wspominaliśmy wcześniej: czynienie dobra jednemu często prowadzi do wyrządzania krzywdy drugiemu.

Prawdopodobnie w odpowiedzi na ten rodzaj problemów Asimov rozszerza listę zasad i dodaje nowe prawo, jeszcze bardziej podstawowe, które nazywa „Prawem Zero” właśnie dlatego, iż ma ono być nadrzędne wobec trzech pozostałych.

0. Robot nie może wyrządzić szkody ludzkości, ani pozwolić na to, aby przez zaniechanie jego działania, ludzkość doznała szkody.

Wprowadzona zostaje tu abstrakcyjna zasada, która nakłada na robota obowiązek myślenia w sposób bardziej uniwersalny. Prawo to zostało wprowadzone w opowiadaniu *Roboty i Imperium*⁸⁾. Jednak nie rozwiązuje ono problemu do końca. Robot Giskard, aby postąpić zgodnie z nim, łamie Pierwsze Prawo, co doprowadza do zniszczenia jego pozytonowego mózgu. Właśnie te sprzeczności między poszczególnymi prawami w sytuacji, gdy muszą zostać skonfrontowane z rzeczywistością, pozwalają zrozumieć, jak ważna jest dla rozwoju etyki robotów twórczość literacka Asimova i innych pisarzy fantastyki naukowej, którzy zgłębili jej tematykę.

Należy pamiętać, iż znaczenie badań Asimova przekracza granice literatury. Japoński rząd na przykład narzuca konstruktorom zbiór zasad bazujących na prawach opracowanych przez rosyjskiego pisarza, właśnie po to aby nie dopuścić do wyrządzenia szkód ludziom. Tak tę wiadomość opisuje włoski dziennik *La Repubblica*: „rząd japoński bardziej niż o to aby roboty były coraz bardziej podobne do wspaniałego Rutgera Hauera z *Bladerunner*, troszczy się o to, aby nie miały one jego siły destrukcji. W świecie, w którym rzeczywistość jest coraz bliższa *science-fiction*, Ministerstwo Gospodarki, Handlu i Przemysłu w Tokio pracuje nad listą zasad bezpieczeństwa dla producentów robotów nowej generacji, które będą gotowe przed końcem roku. Konieczność określenia zasad bezpieczeństwa wynika z obawy, iż coraz bardziej wyspecjalizowane maszyny mogą stać się niebezpieczne dla ludzi. Istotnie, jedna ze stron powieści fantastyczno-naukowej stała się faktem: w 1940 r. Isaac Asimov, autor *Trylogii galaktycznej* i pisarz, który najlepiej zdefiniował relacje między człowiekiem a robotem, ogłosił pierwsze prawo robotyki: „robot nie może skrzywdzić człowieka, ani przez zaniechanie działania dopuścić, aby człowiek doznał krzywdy”. Dyrektywy, nad którymi pracuje japońskie ministerstwo, dążą właśnie do tego, jako że wymagają, by producenci instalowali w swoich robotach odpowiednią liczbę czujników uniemożliwiających im zderzenie się z człowiekiem, a do ich produkcji stosowali miękkie materiały, redukujące skutki takiego ewentualnego

zderzenia. Wśród zasad jest również obowiązek wyposażenia robotów w łatwo dostępne i proste w użyciu wyłączniki. Idea takiego nakazu jest zrozumiała, jeżeli przypomnimy sobie wiele historii fantastyczno-naukowych, w których biedny człowiek nie był w stanie zatrzymać rozszalałego robota”⁹⁾.

Artykuł kładzie też nacisk na fakt, iż we Włoszech, gdzie roboty humanoidalne są mało popularne, japońska inicjatywa wydaje się stać na pograniczu fantastyki naukowej. Tam natomiast, gdzie roboty zadomowiły się już w środowisku człowieka, wprowadzenie odpowiedniego prawa staje się konieczną odpowiedzią na potrzeby obywateli. W Imperium Wschodzącego Słońca, cierpiącym na chroniczny niedobór pracowników sektora opieki społecznej i pielęgnarskiego, przemysł robotów rozwija się w niesamowitym tempie m.in. właśnie po to, aby stworzyć automaty mogące zastąpić człowieka w opiece nad osobami starszymi, chorymi lub kalekimi.

W Japonii problem z osobami do sprawowania opieki nie mógł zostać rozwiązany, podobnie jak to miało miejsce we Włoszech, na skutek masowej imigracji z obszarów ekonomicznie mniej rozwiniętych. Jeśli spojrzeć na warunki demograficzne i geograficzne, można odkryć, że ten azjatycki kraj jest stosunkowo przeludniony. Ponad 127 milionów mieszkańców zaludniających obszar trochę tylko większy od Włoch, gdzie gęstość zaludnienia wynosi 337 mieszkańców na kilometr kwadratowy (w stosunku do 199,9 mieszkańców/km² na Półwyspie Apenińskim). A już nawet Włochy należą do krajów gęsto zaludnionych w stosunku do innych rejonów naszej planety. Tak więc nie może dziwić fakt, iż Japonia woli zwrócić się w stronę urządzeń humanoidalnych zamiast korzystać z pomocy emigrantów i bardzo wspiera rozwój robotyki.

Ostatnie wynalazki techniki japońskiej to: Ri-man – robot, który potrafi zająć się osobami niepełnosprawnymi i Wabian-2 – robot, który potrafi zmieniać wyraz twarzy w zależności od nastroju, jaki chce wyrazić (ten drugi będzie produkowany także we Włoszech). W Japonii obowiązuje już prawo dotyczące produkcji robotów przemysłowych (*Occupation Health and Safety Law*), lecz obecnie odczuwa się potrzebę specjalnej legislacji dotyczącej robotów domowych nowej generacji. Zważywszy że w Imperium Wschodzącego Słońca dochód przemysłu robotyki w 2005 r. wyniósł ponad sześć tysięcy miliardów jenów, skrupulatność, z jaką dąży się do opracowania regulacji etyczno-prawnych produkcji robotów nie może dziwić. Cristina Nadotti podkreśla, że „pod koniec 2004 r. w Japonii w użyciu było ponad 356 tysięcy robotów przemysłowych – najwięcej na świecie. Aby zobrazować przepaść w stosunku do innych krajów, wystarczy porównać dystans, jaki dzieli Tokio od Stanów Zjednoczonych, drugiego państwa na świecie pod względem wykorzystania robotów, z wynikiem „zaledwie” 122 tysięcy robotów”¹⁰⁾. METI (japońskie Ministerstwo Technologii) przewiduje, że do 2020 r. dochody przemysłu robotyki przekroczą dochody przemysłu samochodowego.

⁸⁾ I. Asimov, *Robots and Empire*, Grafton Books, Londyn, 1985.

⁹⁾ C. Nadotti, *Tokyo prepara le regole per i robot: Non danneggino gli esseri umani* [Tokio przygotowuje zasady dla robotów: Niech nie szkodzą ludziom], *La Repubblica*, 29 maja 2006.

¹⁰⁾ ibidem.

3. Kodeks EURON

Istnieją kodeksy etyki robotów alternatywne do kodeksu Asimova, który, jak widzieliśmy, jest trudny do pogodzenia z wojskowym zastosowaniem robotów, i jest źródłem wielu paradoksów. Jeden z takich kodeksów został opracowany przez międzynarodowy zespół naukowców należących do EURON (European Robotics Research Network), dlatego prowizorycznie nazywany jest „Kodeksem EURON”. Zgodnie z Gianmarco Verruggio priorytety są dwa i są ze sobą ściśle powiązane: „Musimy opracować etykę dla naukowców konstruujących roboty oraz sztuczną etykę, którą zostanie zaprogramowana w robotach. Naukowcy muszą zacząć analizować tego rodzaju zagadnienia i badać, czy regulamin lub prawo są potrzebne do tego, by chronić obywateli. Inteligencja robotów będzie coraz większa i, pod pewnymi względami, potężniejsza od inteligencji ludzkiej. Ale pozostanie ona obcą inteligencją. Dlatego też wolałbym dać pierwszeństwo ludziom”¹¹). Odczuwalna jest więc konieczność natychmiastowego zajęcia się problemem, zanim roboty staną się inteligentniejsze, szybsze i silniejsze od człowieka, a ich obecność zdominuje nasze społeczeństwo.

Strategia polegająca na daniu priorytetu człowiekowi jest zgodna z założeniami Asimova, który postrzega roboty jedynie jako maszyny nam podporządkowane i podlegające naszej kontroli. Jednak kodeks EURON nie zajmuje się jeszcze formułowaniem szczegółowych praw lub zasad, koncentrując się jedynie na formułowaniu ogólnych zaleceń, które następnie mogą być użyte jako podstawa bardziej szczegółowych regulacji. Wyróżniono następujące priorytety: producenci muszą tak budować maszyny, aby pozostawały one zawsze pod kontrolą człowieka; nie można dopuścić do nielegalnego wykorzystania maszyn; należy chronić dane uzyskane przez roboty; należy opracować zrozumiałą system identyfikacji maszyn, jak również sposób ich namierzania.

Oto synoptyczny obraz pięciu zaleceń dotyczących etyki robotów:

- **Safety:** *Ensure human control of robot* (robot musi być sterowany przez człowieka)
- **Security:** *Prevent wrong or illegal use* (zapobiegaj niewłaściwemu lub nielegalnemu użyciu robota)
- **Privacy:** *Protect data held by robot* (chronić dane przechowywane przez robota)
- **Traceability:** *Record robot's activity* (monitoruj działania robota)
- **Identifiability:** *Give unique ID to each robot* (nadaj unikalny identyfikator każdemu robotowi).

Innymi słowy są to normy zdroworozsądkowe, mające na celu zapobieganie, zniechęcanie lub zakazywanie ewentualnych niemoralnych lub nielegalnych zastosowań maszyn, bądź też skuteczne reagowanie w razie nagłego „szaleństwa” tychże samych. Mafie, grupy terrorystyczne, zamożni kryminaliści mogliby w rzeczywistości użyć robotów do zabi-

jania, kradzieży, zastraszania, grożenia, szantażu. Robot nie obawia się aresztowania lub kary, miałby znacznie mniej skrupułów psychologicznych i operacyjnych niż morderca będący człowiekiem, a więc byłby zdecydowanie groźniejszy. Wraz z obniżeniem kosztów, coraz większa liczba zлочyńców zaczęłaby powierzać wykonanie swoich kryminalnych zamyśłów maszynom. Oto dlaczego potrzebne byłoby zainwentaryzowanie i skatalogowanie wszystkich robotów w obiegu, ustalając jednocześnie ochronę i odpowiedzialność za ich czyny ze strony ich właścicieli, jak również ustalając sankcje w razie sprzedaży lub zakupu robota niezarejestrowanego w bazie danych. W ten sposób postępujemy już w przypadku środków transportu, które, należy o tym pamiętać, są jedną z głównych przyczyn śmierci w nowoczesnym społeczeństwie. Baza danych o robotach nie rozwiązałaby wszystkich problemów, lecz pomogłaby je znacznie ograniczyć.

Zalecenia EURON-u mogą stanowić dobrą bazę wyjściową, ale wdrożenie tych podstawowych zasad napotyka nielatte do rozwiązania praktyczne problemy. Przede wszystkim należałoby lepiej określić, co oznacza sformułowanie „kontrola ze strony człowieka”. W ogólnym wyobrażeniu robot nie jest tylko zdalnie sterowaną maszyną, lecz maszyną mającą, w większości wypadków, pewną samodzielność w podejmowaniu decyzji (również w obrębie programu lub w określonych specyficznych warunkach). Kontrola może więc zaczynać się od wydawania przez człowieka komend dotyczących każdego wykonywanego przez robota ruchu, aż do sytuacji, w której robot jest zasadniczo „wolny”, ale w każdej chwili może zostać wyłączony przez człowieka. Tak więc kwestia kontroli podlega siłą rzeczy gradacji. Jak ustalić idealny punkt w tej skali? Musimy również postawić sobie pytanie: „kontrola ze strony człowieka” w jakim sensie? Ze strony dowolnego człowieka? Czy przez uprawnionego właściciela robota? A w razie jego braku, kto powinien sprawować tę kontrolę?

Jeśli odpowiedź brzmi: „przez dowolnego człowieka”, to staje się coraz mniej możliwe wykorzystanie maszyn w działaniach wojennych, do ochrony lub do jakiegokolwiek innej działalności (również ekonomicznej), która faworyzuje jednego człowieka kosztem drugiego. Pamiętajmy, że u podstaw gospodarki kapitalistycznej leży rywalizacja oraz że skutki złe dokonanych wyborów mogą okazać się nawet śmiertelne dla poszkodowanego ekonomicznie podmiotu. Oznacza to, że maszyna może zabić człowieka, nie tylko strzelając do niego, lecz również „kradnąc” mu miejsce pracy. Tak więc jeśli *wszyscy ludzie* mieliby władzę nad *wszystkimi robotami*, w jakim celu wojsko lub jakaś wielka korporacja miałaby kupować roboty? Wydaje się więc uzasadnione myślenie, że to właściciele będą mogli swobodnie dysponować swoimi robotami, włączając je i wyłączając wedle uznania i wykorzystując je (mamy taką nadzieję) zgodnie z prawem. Ale nigdy w społeczeństwie bazującym na współzawodnictwie i na kontroli biurokracji nie będziemy mogli mówić o kontroli, w dosłownym tego słowa znaczeniu, ludzi nad robotami¹²). To właśnie

¹¹) E. Habershon i R. Woods, *No sex please, robot, just clean the floor*, The Sunday Times, 18.06.2006.

¹²) Aby poznać spojrzenie futurologiczne na rozwój roboetyki w społeczeństwie bazującym na rywalizacji, zob. Moravec H., *Bodies, Robots, Minds*, 1995: <http://www.frc.ri.cmu.edu>; lub, tego samego autora, *The Universal Robot*, Siemens Review, v. 60/1, January/February 1993: 36–41.

roboty używane do ochrony i nadzoru wywracają do góry nogami ten paradygmat, zważywszy na to, że to właśnie one „kontrolują” ludzi na zlecenie innych ludzi.

Każda technologia sprzyja pewnym grupom ludzi (zazwyczaj tym, którzy nią dysponują), zgodnie z prawem Bacona: *scientia est potentia* (wiedza to władza). Tak więc raczej naiwnym byłoby stawianie sobie pytania, czy dana technologia, łącznie z robotyką, jest dobra czy zła dla całej ludzkości. Ponadto, jeżeli chodzi o zagadnienia związane z *safety*, należy pamiętać, iż „kontrola człowieka” nad robotami niczego nie gwarantuje, jeżeli nie zostaną określone cechy moralne oraz cele konkretnej osoby, która taką kontrolę sprawuje. Mafia i grupy terrorystyczne to też części „ludzkości”.

Szczególnie znaczące na poziomie etyki i przezorności jest zatem pytanie dotyczące roli robotów w utrzymaniu porządku publicznego. Ronald Arkin z GalTech utrzymuje, że istotnie „należy przemyśleć, co mogłoby się stać, jeżeli robotom zostałyby powierzone zadania kontroli tłumu oraz nadane, na przykład, uprawnienia do tłumienia siłą masowych zamieszek”¹³⁾.

W takim wypadku znaleźlibyśmy się poza horyzontem etycznym Praw Asimova, lecz takie zastosowanie byłoby jeszcze uzasadnione przy odpowiedniej interpretacji Kodeksu EURON (gdzie „kontrola ze strony człowieka” mogłaby być rozumiana jako możliwość do zatrzymania maszyn w każdym momencie *przez ich właścicieli*, a nie przez kogokolwiek). Kwestia hipotetycznego tłumienia zamieszek powinna zostać poddana ocenie w świetle prawa obowiązującego w danym kraju. Roboty nie powinny mieć uprawnień do robienia tego, czego nie może robić sama policja.

Podstawowym zagadnieniem pozostaje jednak odpowiedzialność. Jeśli policjant nadużyje władzy w działaniach represyjnych, a więc użyje niepotrzebnie przemocy wobec manifestujących, może zostać zidentyfikowany, a następnie ukarany. Tak samo, jeśli policjant wyrządza krzywdę lub zabija osobę manifestującą, lecz zostanie udowodnione, że działał w obronie własnej (by chronić swoje życie lub swoją nieetykalność osobistą), nie może zostać ukarany lub zostaną wzięte pod uwagę okoliczności łagodzące. Ale w przypadku robota nie byłoby możliwe odwoływanie się do takich okoliczności łagodzących, jako że nie jest on ani żywy, ani świadomy. Zatem jeśli armia robotów zostałaby zatrudniona do publicznych służb porządkowych i przyczyniłaby się do zranienia lub wręcz śmierci człowieka, koniecznym okazałoby się ustalenie osoby odpowiedzialnej i winnego. Jeśli wina zostałaby przypisana producentom lub sprzedawcom, ci najprawdopodobniej *z góry* stwierdziliby, iż wykorzystanie robotów do tych celów było niewłaściwe, nie biorąc na siebie żadnej odpowiedzialności. Jeżeli sąd zdecydowałby, że odpowiedzialność spoczywa na osobie, która wydała polecenie i wysłała roboty do stłumienia zamieszek, prawdopodobnie roboty nigdy więcej nie zostałyby użyte do podobnych celów. Zostałyby wykorzystane do sprawowania nadzoru, ale kontakt fizyczny z człowiekiem należałby jedynie do policjantów. Tak będzie się działo przynajmniej w państwach prawa. Otwarty pozostaje jednak problem etyczny wykorzystania robotów w reżimach totalitarnych. Mogłby również wrócić w krajach demokratycznych, jeżeli uległaby zmianie natura samych robotów i sztucznej inteli-

gencji. Znany futurolog Ian Pearson utrzymuje, że za kilka lat nie będziemy mogli traktować robotów jedynie jak przedmioty: „przewiduję, że do 2020 r. będziemy mieli maszyny świadome”. Czyli za dziesięć lat. „Jeśli wgramy robotom świadomość, staną się androidami. I to będzie ogromną zmianą etyczną”¹⁴⁾.

W takim wypadku same roboty mogłyby upomnieć się o swoje prawa, a akt ich wyłączenia przez człowieka mógłby być postrzegany nawet jako morderstwo. Ale chyba jeszcze za wcześnie na zajmowanie się tym problemem.

¹³⁾ E. Habershon i R. Woods, *No sex please, robot, just clean the floor*, op. cit.

¹⁴⁾ ibidem.

Two Roboethics Approaches: the Problem of Human Control

Abstract: In this article, I briefly present two quite known roboethics approaches: Asimov's "Laws of robotics" and the EURON's "Roboethics roadmap". Then I compare their principles, in the view of some practical problems concerning robotics – especially the use of robots in war and for vigilance tasks. I show that Asimov's approach prohibits in principle violent activities performed by robots, while EURON's five principles of roboethics (*Safety*: Ensure human control of robot; *Security*: Prevent wrong or illegal use; *Privacy*: Protect data held by robot; *Traceability*: Record robot's activity; *Identifiability*: Give unique ID to each robot) rather regulate it. Finally, I show that the principle of safety must be better specified and cannot be seen as a "panacea". It will be always necessary to take into account two aspects: a) the degree of autonomy of robots grows with their sophistication; b) since "humanity" is a nebulous concept, it is necessary to say explicitly which group of humans should exert control over robots and in which circumstances.

Keywords: robot, humanoid, roboethics

dr hab. Riccardo Campa, prof. UJ

Urodzony w 1967 r. w Mantui (Włochy), obecnie wykładowca w Instytucie Socjologii Uniwersytetu Jagiellońskiego w Krakowie. Doktor nauk politycznych (Uniwersytet w Bolonii) oraz filozofii (Uniwersytet Mikołaja Kopernika w Toruniu). Autor czterech książek: *Epistemological Dimensions of Robert Merton's Sociology* (2001), *Il filosofo è nudo* (2001), *Etica della scienza pura* (2007) oraz *Mutare o perire. La sfida del transumanesimo* (2010). Założyciel i przewodniczący Włoskiego Stowarzyszenia Transhumanistów, członek Institute for Ethics and Emerging Technologies. Często publikuje w wydawanym przez włoskich socjalistów *MondOperaio*.

e-mail: riccardo.campa@gmail.com

