

## Kodeksy etyczne robotów bojowych

**Słowa kluczowe:** roboty bojowe, robolucja, autonomia, system etyczny, prawa robotyki

**Keywords:** robolution, autonomy, ethical system, robotics laws

### Wprowadzenie

Roboty bojowe są jednym z najważniejszych kierunków rozwoju techniki wojskowej. Proliferacja tzw. warbotów (określenie powstałe z połączenia słów wojna – ang. war – i robot) prowadzić ma do kolejnej rewolucji militarnej, tzw. robolucji<sup>1</sup>. Ten chwytliwy termin (połączenie robotyzacji i rewolucji) opiera się na założeniu, że transformacja sposobu prowadzenia wojny, wynikająca z zastosowania robotów, nie ograniczy się do sfery techniki. Co prawda zmiana technologiczna stoi u podstaw nadchodzącej rewolucji, ale ma ona doprowadzić do przekształcenia pozostałych kluczowych sfer wojny – jej doktryny oraz organizacji. Pojawiają się również twierdzenia, że skutki robotyzacji będą znacznie bardziej dalekosiężne, zmieniając nie tylko charakter, ale i naturę wojny<sup>2</sup>.

---

1 A. Windeck, *Preface*, w: *Robots on the Battlefield, Contemporary Issues and Implications for the Future*, R. Doaré i in. (red.), Fort Leavenworth 2014, s. V.

2 D. Danet, J.-P. Hanon, *Digitization and Robotization of the Battlefield: Evolution or*

Optymistyczne wizje, koncentrujące się na perspektywie prowadzenia wojen pozbawionych ryzyka dla żołnierzy<sup>3</sup>, miesza się z ostrzeżeniami przed trudnymi do przewidzenia konsekwencjami eliminacji człowieka z pętli decyzyjnej. Widzimy te obawy choćby w działalności Campaign to Stop Killer Robots, powstałej w 2012 roku koalicji organizacji pozarządowych nawołujących do wprowadzenia zakazu broni autonomicznej<sup>4</sup>. Dostzegamy je także w rekomendacjach Christopha Heynsa, specjalnego sprawozdawcy ONZ w sprawie selektywnej eliminacji (*targeted killings* – likwidowanie osób podejrzanych o terroryzm, realizowane przede wszystkim przy pomocy samolotów bezzałogowych, tzw. dronów<sup>5</sup>), który nawoływał, by państwa ustanowiły swoje własne moratoria na rozwój śmiertelnych robotów autonomicznych<sup>6</sup>.

## 1. Decyzyjność robotów

Tym, co czyni roboty bojowe źródłem tak wielu obaw jest zdolność do podejmowania decyzji co do życia i śmierci. Przeróża nie to, że roboty bojowe będą w stanie zabijać – tę zdolność posiadają już najprostsze systemy automatyczne, z minami na czele. Roboty będą mogły jednak decydować – zabić czy nie zabić<sup>7</sup>. Używam czasu przyszłego, gdyż w pełni autonomiczne

---

*Robolution?*, w: *Robots on the Battlefield...*, dz. cyt., s. XIII.

3 P.M. Asaro, *How Just Could a Robot War Be?*, w: *Current Issues in Computing and Philosophy. Frontiers in Artificial Intelligence and Applications*, A. Briggie, K. Waelbers, P. Brey (red.), Amsterdam 2008, s. 50-64.

4 *Campaign to Stop Killer Robots*, <https://www.stopkillerrobots.org/>, [23.02.2016].

5 R. Kopeć, *Dyplomacja dronów*, „Kultura i Polityka”, 2015, nr 17, s. 65-83.

6 *A call for moratorium on the development and use of lethal autonomous robots*, United Nations Human Rights, Office of the High Commissioner, <http://www.ohchr.org/EN/NewsEvents/Pages/Acallformoratoriumonthedevelopmentrobots.aspx>, [25.02.2016].

7 A. Krishan, *Killer Robots. Legality and Ethicality of Autonomous Weapons*, Farnham 2009, s. 33.

systemy bojowe są kwestią przyszłości. Przez systemy w pełni autonomiczne rozumiem systemy zdolne do autonomicznej selekcji i rażenia celów, a także posiadające możliwość autonomicznego poruszania się, tzw. roboty *free-ranging*, czyli mówiąc kolokwialnie „z wolnego wybiegu”. Systemy spełniające zasadnicze kryteria autonomii, jednakże mające charakter defensywny i niezdolne do autonomicznego poruszania się, już jednak istnieją i są wykorzystywane przez siły zbrojne. Chodzi przede wszystkim o kompleksowe systemy przeciwlotnicze i przeciwraкетowe, a także o tzw. roboty-strażników.

Możliwość samodzielnej selekcji celów oraz podejmowania decyzji o ich rażeniu przenosi roboty na inny poziom ontologiczny – na poziom bytów zdolnych podejmować decyzje, które mają znaczenie moralne. To właśnie zdolność do decydowania, a nie sama potencjalna śmiertelność, stanowi powód największych obaw. Nie chodzi bynajmniej o to, że roboty będą bardziej śmiertelne niż np. miny. Być może tak będzie, być może nie – poza tym, jaką miarę śmiertelności przyjąć?

Thomas K. Adams twierdzi, że „przerzucanie” decyzji na roboty stanowi element szerszego trendu w kierunku ograniczenia ludzkiej decyzyjności na wojnie<sup>8</sup>. Następuje bowiem komplikowanie środowiska informatycznego pola bitwy, chociażby poprzez zdolność do pozyskiwania informacji z wielu źródeł. Człowiek nie jest w stanie przetworzyć takiej ilości informacji, więc ucieka się do pomocy systemów automatyzacji (jeszcze nie autonomii) procesu dowodzenia. Zazwyczaj pomoc ta polega na wsparciu w podejmowaniu decyzji (np. syntezie informacji i proponowaniu rozwiązań), ale stanowi to pierwszy krok w kierunku eliminacji człowieka z pętli decyzyjnej. Tendencja w kierunku zwiększenia

---

8 T.K. Adams, *Future Warfare and the Decline of Human Decisionmaking*, „Parameters”, 2001, nr 31:4, s. 57-71.

liczby dostępnych informacji skutkuje brakiem możliwości zgregowania i syntezy ich przez ludzki umysł. Komputer jawi się jako idealne rozwiązanie tego problemu, gdyż to właśnie przetwarzanie informacji jest domeną komputerów i polem, na którym w sposób zasadniczy przewyższają one ludzi. Dodatkową presję tworzy konieczność ograniczenia czasu na podjęcie decyzji, zwłaszcza odkąd szeroko rozumiana prędkość stała się jednym z decydujących czynników na polu walki<sup>9</sup> (np. środki rażenia charakteryzują się tak dużą prędkością, że pozostawiają mało czasu na reakcję). Decyzję wypracowują więc automaty. Jak dotychczas, człowiek zazwyczaj ma możliwość ich zatwierdzenia lub zablokowania, ale często ta możliwość jest iluzoryczna, gdyż człowiek wie, że maszyna jest w procesie podejmowania decyzji bardziej kompetentna<sup>10</sup>.

## 2. Broń czy żołnierze?

By prowadzić rozważania na temat roli robotów na polu walki i skutków tej obecności, należy zdefiniować to pojęcie. Robot charakteryzuje się następującymi cechami:

- a) zdolność manipulacji symbolami – robot jest więc po części komputerem (czyli charakteryzuje go rodzaj sztucznej inteligencji),
- b) zdolność manipulowania obiektami fizycznymi – robot jest więc po części maszyną,

---

9 Ł. Kamieński, *Technologia i wojna przyszłości. Wokół nuklearnej i informacyjnej rewolucji w sprawach wojskowych*, Kraków 2009, s. 260-262.

10 R. Crootof, *The Killer Robots Are Here, Legal and Policy Implications*, „Cardozo Law Review”, 2015, nr 36:1837, s. 1849.

- c) elastyczność – zdolność stosowania różnych sposobów manipulacji symbolami i manipulacji obiektami fizycznymi (to różni go od różnych systemów automatycznych, zaprogramowanych tylko na realizację z góry określonego działania).

Istnieją dwa spojrzenia na rolę robotów na polu walki. Pierwsze zakłada, że są one przedłużeniem człowieka (zgodnie z tezą McLuhana twierdzącego, że każda technologia jest tylko przedłużeniem ludzkiego ciała<sup>11</sup>). W tym sensie roboty są kolejnym krokiem w kierunku realizacji dwóch współzależnych trendów w zakresie ewolucji broni. Pierwszy to dążenie do zwiększenia dystansu na polu walki w celu osiągnięcia takiego stanu umożliwiającego rażenie przeciwnika z odległości, która czyni nas odpornym na jego przeciwdziałanie. Ten trend jednak dotychczas napotykał na ograniczenie – im większy dystans, tym mniejsza precyzja<sup>12</sup>. Trzeba więc albo zwiększyć siłę rażenia (ostatecznym i krańcowym wyrazem tej tendencji jest broń nuklearna), albo zaimplementować różne rodzaje precyzyjnego kierowania. To jest właśnie druga tendencja – tendencja do budowy tzw. broni inteligentnej (*smart weapon*). To chwytliwe, ale mało precyzyjne określenie. Tak naprawdę chodzi o różne formy precyzyjnego kierowania – od zdalnego sterowania do samonaprowadzania (tzw. broń *fire and forget* – odpal i zapomnij). Samonaprowadzanie to jednak zazwyczaj działanie na prostej zasadzie bodziec-reakcja. Przykładem jest kierowanie rakiety naprowadzanej na podczerwień w stronę źródła ciepła. Trudno tu mówić o inteligencji. Inteligencja zakłada bowiem zdolność do abstrakcyjnego myślenia. Tu pojawia się drugie spojrzenie na rolę robotów

---

11 M. McLuhan, *Understanding Media: The Extension of Man*, Nowy Jork 1964, s. 152.

12 P. Scharre, *Robotics on the Battlefield. Part I: Range, Persistence and Daring*, Waszyngton 2014, s. 19.

na polu walki, postrzegające roboty jako niezależnych agentów, jako byty niemal równoważne żołnierzom.

Tego typu roboty – autonomiczne, zdolne do samodzielnego podejmowania decyzji i „inteligentne” (choć kwestia definiowania inteligencji i możliwości zbudowania jej sztucznego odpowiednika jest przedmiotem niekończących się sporów) – będą znacząco różniły się od innych rodzajów broni inteligentnej. Różnica nie będzie miała charakteru ilościowego (większe natężenie danej cechy, większe możliwości w tym samym zakresie), ale jakościowy. W pewnym sensie dopiero autonomiczne roboty bojowe będą zasługiwały na określenie „broń inteligentna”. Co więcej, można wręcz zastanawiać się, czy będzie je można nazywać bronią, czy też staną się one kolejnym rodzajem „bytów pola walki” usytuowanych co najmniej w połowie drogi między bronią a żołnierzem, a być może nawet w pewnym sensie równoważnym żołnierzom. Autonomiczny robot bojowy nie będzie więc tylko „maszyną do zabijania” (w zasadzie każda broń nią jest), ale stanie się czymś więcej – panem życia i śmierci. W tym sensie roboty przestaną być bronią zwiększającą możliwości żołnierzy (tak postrzega to pierwsza szkoła), a same stają się quasi-żołnierzami (tak widzi je druga szkoła)<sup>13</sup>.

---

13 D.F. Bigelow, *Fast Forward to the Robot Dilemma*, „Armed Forces Journal”, 2007, nr 11, s. 16-22.

### 3. Etyka robotów

W twórczości *science-fiction* występuje wyraźna tendencja do nadawania robotom ludzkich kształtów, czyli tworzenia tzw. robotów humanoidalnych. W dużej mierze stan taki przekłada się na specyficzne relacje na linii ludzie-roboty, chociażby poprzez proces antropomorfizacji, czyli przypisywania innym bytom cech ludzkich. Tworzenie robotów humanoidalnych ma też uzasadnienie funkcjonalne. Po pierwsze, w pewnych zadaniach rozmieszczenie sensorów i efektorów wzorowanie na ludziach jest optymalne. Po drugie, roboty często operują w przestrzeni stworzonej z myślą o ludziach, tak więc humanoidalne maszyny będą się tam najlepiej odnajdywały (chodzi o rozmiar czy pewne cechy mobilności, np. zdolność chodzenia po schodach).

Dla roli robotów na polu walki decydujący jest jednak nie kształt czy budowa mechaniczna, lecz możliwości kognitywne. Decyzje dotyczące życia i śmierci to decyzje mające konsekwencje moralne i jako takie powinny opierać się na podstawach moralnych. W przypadku robotów bojowych nie sposób oczywiście uciec od wartości mających zasadnicze znaczenie, czyli wartości użytecznych (przede wszystkim militarnej użyteczności), ale decyzje o życiu i śmierci powinny również bazować na kodeksie etycznym.

W zasadzie należy zadać pytanie – dlaczego? Dlaczego w ogóle roboty mają mieć wbudowane systemy etyczne? Czy działanie tylko i wyłącznie w oparciu o kryterium militarnej skuteczności nie jest wystarczające? Twierdzę, że nie. I to zarówno z przyczyn normatywnych (sytuujących się w sferze wartości), jak i instrumentalnych (odnoszących się do sfery możliwości realizacji danego zadania). Po pierwsze chodzi o sposób, w jaki państwa szeroko pojętego Zachodu (bardziej pod względem kulturowym

niż geograficznym) prowadzą obecnie wojny. Koncentracja na państwach zachodnich wynika z faktu, że w dalszym ciągu są one liderami technologicznymi w zakresie uzbrojenia oraz sprzętu wojskowego i wytyczają kierunki jego rozwoju. Kluczowa rola technologii jest również sztandarową cechą zachodniego sposobu prowadzenia wojny<sup>14</sup>. Nawet jeśli armie należące do innych kręgów cywilizacyjnych (np. chińska) podążać będą w kierunku robotyzacji, zapewne nie będzie to miało aż takiego znaczenia, jak w przypadku Zachodu, gdzie robotyzacja stać się może osią rozwoju techniki wojskowej. Zachodni sposób prowadzenia wojen tradycyjnie opiera się na modelu kapitałochłonnym (co skutkuje naciskiem położonym na innowacyjność), a nie „ludzkochłonnym”. Obecnie tendencję tę pogłębia również przemożna chęć uniknięcia strat własnych, która od czasu „Pustynnej Burzy” stała się centralnym imperatywem prowadzenia wojny przez kraje rozwiniętego Zachodu.

Zachód obecnie prowadzi wojny w duchu politycznej poprawności. Nie do zaakceptowania są wysokie straty wśród ludności cywilnej na terenach objętych działaniami zbrojnymi. Przeciwnicy potrafią to zręcznie wykorzystywać, prowokując do ostrzału obiektów cywilnych, a następnie nagłaśniając takie przypadki (patrz: działania Hamasu<sup>15</sup>). Roboty bojowe nie mogą więc działać na ślepo, lecz muszą być zdolne do starannej selekcji celów. Co więcej, zdolność ta obejmować ma najtrudniejsze wyzwanie współczesnych nieliniarnych konfliktów zbrojnych, czyli rozróżnienie między walczącymi (którzy często nie są wcale członkami regularnych sił zbrojnych), a ludnością cywilną. Z tym zadaniem

---

14 G. Parker, *Historia sztuki wojennej. Od starożytności do czasów współczesnych*, Warszawa 2008, s. 14.

15 S. Erlanger, *A Gaza War Full of Traps and Trickery*, „New York Times”, [http://www.nytimes.com/2009/01/11/world/middleeast/11hamas.html?partner=permalink&exprod=permalink&\\_r=0](http://www.nytimes.com/2009/01/11/world/middleeast/11hamas.html?partner=permalink&exprod=permalink&_r=0), [28.02.2016].



nie radzą sobie nawet ludzie, dlatego często słyszymy o ofiarach cywilnych będących skutkiem pomyłki<sup>16</sup>.

Co więcej, konieczność zastosowania kodeksu etycznego to nie tylko pochodna systemu wartości, w imię których Zachód prowadzi wojny, ale także rezultat swoiście rozumianego utilitaryzmu. Państwa zachodnie prowadzą bowiem często tzw. wojny o serca i umysły (czego przykładem są operacje w Iraku i Afganistanie). Są to wojny, w których stawką jest przeciągnięcie na swoją stronę społeczeństw w państwach, w których się ona toczy, a nie „stłamszenie” tych społeczeństw i zmuszenie ich do bezwarunkowej kapitulacji (nie mówiąc już o dążeniu do anihilacji społeczeństw). Dlatego mowa o militarnej politycznej poprawności (bez wartościowania tego pojęcia): państwa zachodnie nigdy nie przyznają, że prowadzą walkę przeciwko danej społeczności, ale twierdzą, że walczą tylko przeciwko pewnym jednostkom, a nawet, że czynią to w imię tej społeczności, by przynieść jej pomyślność i dobrobyt (często utożsamiane z implementacją systemu społeczno-politycznego na modłę zachodnią). Zdolność do rażenia celów, ale bez możliwości ich precyzyjnej selekcji właśnie z uwzględnieniem kryteriów etycznych, w przypadku wojny „o serca i umysły”, pozostanie więc bezwartościowa.

#### 4. Dwa podejścia

Dwa zasadnicze podejścia do budowania systemów etycznych znajdują odzwierciedlenie w podejściu do tworzenia architektury etycznej autonomicznych robotów bojowych. Spojrzenie konsekwencjonalistyczne (teleologiczne), którego ojcem jest Jeremy Bentham, a czołowym nurtem utilitaryzm zakłada, że do-

---

16 J. Griffiths, *Collateral damage: A brief history of U.S. mistakes at war*, CNN, <http://edition.cnn.com/2015/10/06/middleeast/us-collateral-damage-history/>, [27.02.2016].

bre jest to, co przynosi dobre rezultaty. Działanie etyczne to więc takie działanie, które w danym momencie jest postrzegane przez jednostkę jako prowadzące w sposób najbardziej prawdopodobny do najlepszych konsekwencji w przyszłości<sup>17</sup>.

Robot działający etycznie powinien więc w każdej sytuacji wymagającej podjęcia decyzji w sferze moralności zrealizować następujące kroki: dokonać oceny sytuacji, opracować listę potencjalnych sposobów działania, każdemu z nich przyporządkować możliwe przyszłe konsekwencji, wybrać tę opcję, która przyniesie „najwięcej dobra”. Brzmi prosto, ale jednocześnie obrazuje ogrom potencjalnych trudności w realizacji systemu „etycznego sterowania” robota. Pomijając oczywisty fakt, że robot może mylić się w swoich prognozach, zasadnicze pytanie brzmi: jak robot ma mierzyć „dobro”. Z jakiego punktu widzenia ma być ono realizowane: danej osoby, grupy osób, całej ludzkości? W idealnym utylitarystycznym dobro każdej osoby jest traktowane równo, a idealne działanie utylitarne powinno opierać się na zasadzie maksymalizowania ogólnego dobra dla ogółu (ponownie możemy zastanawiać się, co kryje się pod pojęciem ogółu – ludzkość, planeta, wszechświat?).

Drugie podejście odnosi się do deontologii. Deontologia – której rodzajem jest kantyzm – zakłada zupełnie inne spojrzenie: oceniamy działanie samo przez się, nie myślimy o konsekwencjach<sup>18</sup>. Działanie jest z natury dobre albo z natury złe. Najlepszym przykładem podejścia deontologicznego jest Dekalog. Nie ma tam mowy o rozważaniu konsekwencji – po prostu wskazane zostały czyny, które uznane zostały za złe i jako takie są zakazane z punktu widzenia danego systemu etycznego.

---

17 J. Gips, *Towards the ethical robot*, w: *Machine Ethics*, M. Anderson, S.L. Anderson (red.), Cambridge 2011, s. 244-247.

18 Tamże, s. 247-249.

Podejście deontologiczne opiera się na istnieniu zestawu reguł, które są dane. Reguły te mogą pochodzić od Boga (jak w przypadku dekalogu), z rozumnej woli człowieka (imperatyw kategoryczny Kanta), czy zapisów międzynarodowego prawa humanitarnego (które jawią się jako najlepsza baza do budowy kodeksu etycznego autonomicznych systemów bojowych). Katalog reguł powinien zawierać listę działań obligatoryjnych, dozwolonych i zabronionych.

## 5. Porządek reguł

W rozważaniach na temat zagadnień związanych z budowaniem kodeksów etycznych robotów z pomocą przychodzi nam twórczość *science-fiction*. Może to budzić pewne kontrowersje dotyczące wartości naukowej tego typu rozważań, ale moim zdaniem twórczość tego rodzaju może być dla nich użyteczną bazą. Wiele przykładów na wzajemne związki nauki i twórczości *science-fiction* przywołuje Łukasz Kamieński (właśnie z myślą o obronie tego typu materiałów jako źródła naukowych dociekań<sup>19</sup>) – dość powiedzieć, że Joseph Engelberg, twórca maszyny Unimate, uznawanej za pierwszy robot przemysłowy, zainspirował się lekturą opowiadania Isaaca Asimova *Ja, robot*.

Prawa Asimowa, ojca literatury robotycznej w dzisiejszym znaczeniu tego słowa (kreującego roboty jako istoty elektromechaniczne, a nie biologiczne, jak Karel Čapek, twórca pojęcia „robot”), do dziś stanowią najbardziej znany przykład kodeksu etycznego robotów. Co prawda prawa te dotyczą robotów ogólnie, a nie robotów bojowych, a co więcej sposób ich sformułowania w zasadzie wyklucza militarne zastosowanie robotów,

---

19 Ł. Kamieński, *Nowy wspaniały żołnierz. Rewolucja biotechnologiczna im wojna w XXI wieku*, Kraków 2014, s. 74-84.

należy je poddać analizie w celu identyfikacji wyzwań w zakresie tworzenia zespołu reguł mającego kierować działaniem robota z punktu widzenia wymagań moralnych.

Prawa robotyki Isaaca Asimova, sformułowane w opowiadaniu *Zabawa w berka (Runaround)* z 1942 roku przedstawiają się następująco<sup>20</sup>:

1. Robot nie może wyrządzić żadnej krzywdy człowiekowi, ani poprzez brak czynnej reakcji dopuścić, aby człowiekowi stała się jakakolwiek krzywda;
2. Robot musi wykonywać każdy rozkaz człowieka pod warunkiem, że rozkaz taki nie koliduje z Prawem Pierwszym;
3. Robot musi ochraniać własny mechanizm pod warunkiem, że nie jest to sprzeczne z Prawem Pierwszym i Drugim.

Zasadniczą cechą odróżniającą prawa Asimova od np. dekalogu jest hierarchizacja reguł. Każdy system reguł jest bowiem narażony na wewnętrzny konflikt. Prawa robotów muszą tworzyć strukturę hierarchiczną. W przypadku praw Asimova najważniejsze jest nieszkodzenie ludziom, następnie służenie ludziom (rola niewolnika), a na końcu własny interes robota (instynkt przetrwania – sam instynkt każe mu bronić się przed warunkami środowiskowymi, atakiem innego robota czy atakiem człowieka, ale tylko w sytuacji, gdy robot nie jest ograniczony przez prawa wyższego rzędu).

### **Potencjalne problemy**

Analiza praw robotyki Asimova pozwala wskazać kilka potencjalnych problemów, które muszą być rozwiązane, jeśli kodeksy etyczne oparte na zestawie reguł mają stać się elementem

---

20 I. Asimov, *Zabawa w berka*, w: I. Asimov, *Ja, Robot*, Bydgoszcz 1993, s. 52.

systemu sterowania autonomicznych robotów bojowych. Problemy te możemy ująć w postaci następującego katalogu:

1. Dwie osoby o równym statusie wydadzą robotowi sprzeczne rozkazy.
2. Prawa robotyki znajdują się ze sobą w sprzeczności, mimo wprowadzonej hierarchii. Klasycznym przykładem, często przytaczanym do zobrazowania tego dylematu jest sytuacja, gdy robot musiałby skrzywdzić jednego człowieka, by uratować drugiego (np. jeśli jeden człowiek chce zabić drugiego i nie można go powstrzymać inaczej niż przez zabicie go<sup>21</sup>). Cokolwiek uczyni robot, naruszy pierwsze prawo.
3. Problem z interpretacją. Co to znaczy „krzywdzić” albo „uratować”? Te pojęcia wymagają interpretacji<sup>22</sup>. Czy znaczenie słowa „krzywda” może zawierać w sobie krzywdę psychiczną? Inny potencjalny problem – jak robot interpretować ma określenie „człowiek”? Czy w przypadku robotów bojowych pojęcie to obejmuje całą ludzkość, czy może tylko określone grupy ludzi? Kolejny problem: jak robot ma niezawodnie rozpoznawać, że dana sytuacja wymaga aplikacji danej reguły? Do tego potrzebne jest zrozumienie kontekstu, często wymagające dogłębnej znajomości ludzkiej psychologii i swego rodzaju „zdrowego rozsądku”<sup>23</sup>.
4. Problem wypracowania decyzji w niejednoznacznych sytuacjach. Niemożność wypracowania decyzji może spowodować, że robot znajdzie się w sytuacji impasu (*deadlock*),

---

21 R. Campa, *Kodeksy etyczne robotów: zagadnienie kontroli sprawowanej przez człowieka*, „Pomiary Automatyka Robotyka”, 2001, nr 3, s. 67.

22 R. Clarke, *Asimov's Laws of Robotics. Implications for Information Technology, w: Machine...*, poz. cyt., s. 260.

23 N.E. Sharkey, *The inevitability of autonomous robot warfare*, „International Review of the Red Cross”, 2012, nr 94(886), s. 788-789.

z którego sam nie będzie w stanie wyjść. Przykładowo, w sytuacji przywołanej w punkcie 2, robot musi dokonać osądu, które życie jest więcej warte. W jaki sposób ma jednak ocenić, kto jest dobry, a kto zły albo kto jest bardziej użyteczny dla społeczeństwa? Robot może jedynie rozstrzygnąć, jeśli ma wystarczający zasób informacji, kto jest stroną atakującą, a kto broniącą się. To jednak może nie wystarczyć.

Asimov w książce *Roboty i imperium (Robots and Empire)* z 1985 roku uzupełnił stworzone przez siebie reguły, dodając prawo zerowe, nadrzędne wobec trzech pozostałych. Zakłada ono, że:<sup>24</sup>

0. Robot nie może skrzywdzić ludzkości lub poprzez zaniechanie działania doprowadzić do uszczerbku dla ludzkości.

Prawo to pozwala więc skrzywdzić człowieka, jeśli ma to uratować ludzkość. Możemy wyobrazić sobie konieczność likwidacji nosiciela groźnego wirusa, który w innym wypadku wymknie się spod kontroli. Ludzkość jest jednak bytem znacznie bardziej abstrakcyjnym od człowieka. Także przewidywanie rozwoju sytuacji (jaki wpływ będzie miało działanie robota) na taki abstrakcyjny byt, jest zdecydowanie trudniejsze niż w przypadku człowieka, gdzie i tak jest to niezmiernie trudne.

Dodanie tego prawa może też spowodować, że roboty będą zmuszone posługiwać się znacznie bardziej abstrakcyjnymi, ogólnymi kategoriami. Będą robić to, co uznają za słuszne z punktu widzenia abstrakcyjnego dobra (w przypadku zerowego prawa Asimova – całej ludzkości, w przypadku robotów bojowych – np. całej operacji). Można więc mówić o swego rodzaju paternalizmie robotów. Będą one w imieniu ludzi robiły to, co same

---

24 I. Asimov, *Robots and Empire*, s. 285, <http://www.ebooktrove.com/Asimov,%20Isaac/Asimov,%20Isaac%20-%20Robot%2006%20-%20Robots%20&%20Empire.pdf>, [27.02.2016].

uznają za dobre dla ludzi. Może to być pierwszy krok prowadzący do wymknięcia się robotów spod ludzkiej kontroli<sup>25</sup>.

Innym problemem mogącym pojawić się szczególnie w warunkach pola walki jest potrzeba uniknięcia sprzeczności w zakresie rozkazów, które trafiają do robota. Nie chodzi tylko o rozkazy wydawane przez ludzi (jak to wspomniano w punkcie 1), ale przez inne roboty. Jednym z wyzwań robotyki pola walki jest bowiem stworzenie takiego systemu sterowania, by wiele niezależnie funkcjonujących robotów autonomicznych realizowało jeden wspólny cel operacyjny. Należy więc stworzyć strukturę organizacyjną grupy robotów. Może ona mieć charakter hierarchiczny, tak by jeden robot, usytuowany wyżej w hierarchii, mógł wydawać rozkazy robotowi podległemu. Pojawiają się jednakże bardziej zaawansowane koncepcje, szukające inspiracji nie w świecie organizacji stworzonych przez człowieka (np. organizacji militarnych), ale w przyrodzie. Taką modną koncepcją jest oparcie struktury organizacyjnej grupy robotów na strukturze roju, charakterystycznej dla owadów<sup>26</sup>. Ma to pozwolić na lepsze wykorzystanie potencjału rozproszonych, ale współdziałających ze sobą elementów, bez konieczności bezpośredniego sterowania nimi przez ludzkich operatorów.

Na wyzwanie stworzenia kodeksu etycznego obejmującego również kwestie podporządkowania i podległości odpowiadają rozszerzone prawa robotyki Rogera Clarke'. Obejmują one następujące reguły<sup>27</sup>:

Meta-prawo: Robot nie może działać, jeśli jego działanie nie jest w zgodzie z prawami robotyki,

---

25 R. Clarke, *Asimov's Laws...*, poz. cyt., s. 270.

26 P. Scharre, *Robotics on the Battlefield Part II: The Coming Swarm*, Waszyngton 2014, s. 24.

27 R. Clarke, *Asimov's Laws...*, poz. cyt., s. 275.

Prawo 0: Robot nie może wyrządzić krzywdy rodzajowi ludzkiemu, ani – przez zaniechanie działania – dopuścić, aby rodzaj ludzki doznał krzywdy,

Prawo 1: Robot nie może skrzywdzić człowieka, ani przez zaniechanie dopuścić, aby człowiek doznał krzywdy, pod warunkiem, że nie stoi to w sprzeczności z prawami wyższej rangi,

Prawo 2: Robot musi być posłuszny rozkazom człowieka, chyba że stoi to w sprzeczności z prawami wyższej rangi,

Robot musi być posłuszny rozkazom robotów stojących wyżej w hierarchii, chyba że stoi to w sprzeczności z prawami wyższej rangi,

Prawo 3: Robot musi chronić roboty stojące wyżej w hierarchii, chyba że stoi to w sprzeczności z prawami wyższej rangi,

Robot musi chronić sam siebie, chyba że stoi to w sprzeczności z prawami wyższej rangi,

Prawo 4: Robot musi wykonywać zadania, do których został zaprogramowany, chyba że stoi to w sprzeczności z prawami wyższej rangi,

Prawo Prokreacji: Robot nie może brać udziału w projektowaniu i wytwarzaniu robotów, jeśli takie działanie nie jest w zgodzie z prawami robotyki.



Ten zestaw reguł zwraca uwagę na potrzebę ustalenia hierarchii robotów. Otwarte jednak pozostaje pytanie, czy ustanowienie zespołu reguł kierującego działaniem robota w kompleksowym, nieprzewidywalnym środowisku jest w ogólnie możliwe. Z jednej strony reguły te są niejednoznaczne i trudne w interpretacji. Zachodzi więc potrzeba ich uproszczenia, opisania pożądanego i niepożądanego działania w sposób prostszy, mniej abstrakcyjny, bardziej zrozumiały dla robota. Pojawia się jednak pokusa opisania każdej potencjalnej sytuacji, co oczywiście jest niemożliwe. Zachodzi więc sprzeczność pomiędzy dążeniem do rozbudowania i uproszczenia zespołu reguł.

## 6. Wnioski

Tendencja w kierunku robotyzacji wojny budzi wiele obaw z powodu możliwości oddania w ręce robotów decyzji dotyczących życia i śmierci (chodzi dokładnie o decyzje w sprawie selekcji i rażenia celów). Wyposażone w tego typu prerogatywy roboty staną się czymś więcej niż bronią, będą quasi-żołnierzami, co prowokuje oczywiste pytania o odpowiedzialność w przypadku pogwałcenia prawa wojennego. Podejmowanie tego typu decyzji przez maszyny nadaje im również wymiar moralny. W system sterowania robota musi być więc wbudowany system etyczny. Ocena działania robota przez pryzmat tylko i wyłącznie militarnej skuteczności jest niewystarczająca. Ma to szczególne znaczenie w przypadku państw zachodnich (notabene produjących w rozwoju robotów bojowych): po pierwsze, włączenie czynnika etycznego współgra z poglądami na temat metod prowadzenia wojny dominującymi w społeczeństwach Zachodu, po drugie, zaniedbywanie zasad etycznych w działaniach bojowych utrudnia wygranie wojny „o serca i umysły”.

Otwarte pozostaje pytanie, w jaki sposób tego rodzaju system etyczny ma być zaimplementowany w system sterowania robota. Spojrzenie konsekwencjonalistyczne zakłada, że działanie należy oceniać przez pryzmat jego konsekwencji, natomiast spojrzenie odwołujące się do etyki deontologicznej optuje za wbudowaniem zespołu reguł (zawierających katalog działań obligatoryjnych, dozwolonych i zabronionych). Najbardziej znanym, pochodzącym z literatury *science-fiction*, przykładem kodeksu etycznego robotów są prawa robotyki Asimova. Choć nie odnoszą się one wprost do robotów bojowych, a wręcz zabraniają robotom zwracać się przeciwko człowiekowi, stanowią ciekawą podstawę do analizy możliwości budowy tego rodzaju kodeksu. Ustanowienie tego rodzaju kodeksu napotyka na szereg przeszkód odnoszących się m.in. do konieczności unikania sprzeczności, interpretacji skomplikowanych sytuacji (w tym przyporządkowania reguł do sytuacji), czy wypracowania decyzji w sytuacjach niejednoznacznych z etycznego punktu widzenia. Droga do zbudowania systemu sterującego dla robotów bojowych uwzględniającego kryteria etyczne, a przy tym niezaniebującego wymagań z zakresie skuteczności bojowej, wydaje się więc daleka.

## Abstrakt

Zastosowanie robotów w działaniach bojowych przynieść może rewolucyjne zmiany w zakresie techniki, doktryny i organizacji struktur militarnych. Wprowadzenie do uzbrojenia autonomicznych robotów bojowych postrzegane jest często jako otwarcie puszką Pandory, która w bliżej nieokreślonej przyszłości doprowadzić może do tragicznych skutków, nawet do zagłady ludzkości. Obawy te podsycane są przez wytwory kultury (przede wszystkim literaturę oraz film) spod znaku *science-fiction*. Już w sztuce Karela Čapka *R.U.R. (Rossumovi Univerzální Roboti) – Uniwersalne roboty Rossuma* z roku 1921, która wprowadziła do języka nauki i kultury słowo „robot”, roboty zwracały się przeciwko ludziom.

W bliższej perspektywie jednak wyzwaniem jest przede wszystkim etyczny wymiar przeniesienia decyzji dotyczącej życia i śmierci na maszyny. Z jednej strony robotyzacja postrzegana jest jako indyktor tzw. wojny post-ludzkiej (*post-human warfare*), z drugiej być może jest ona naturalnym etapem ewolucji wojen w kierunku zwiększenia dystansu między walczącymi, z technologią traktowaną zgodnie z duchem Marshalla McLuhana, który widział w niej tylko protezę przedłużającą ludzkie ciało. Tekst podejmuje problematykę kontrowersji związanych z budową „komputerowej moralności”, odpowiadając na pytania, jakie założenia towarzyszą temu przedsięwzięciu i jakimi drogami próbuje się je realizować. Koncentruje się przede wszystkim na wyzwaniach związanych z opracowaniem kodeksu etycznego robotów bojowych, przyjmując jako punkt wyjścia prawa robotyki Asimova.

## Abstract

The use of robots in warfare could bring the revolutionary changes in technology, doctrine and military organization. Introducing of autonomous combat robots is often considered as opening the Pandora's box. It may lead to tragic results including human extinction. This anxiety is fueled by science fiction literature and movies. It is enough to mention 1921 Karel Čapek's play *R.U.R. (Rossum's Universal Robots)*, where the term “robot” come from – the robots eventually had stood out there against the humans.

However, in the short term, the main challenge is the ethical dimension of transferring life and death decision to machines. On the one hand, roboti-

sation is perceived as an indicator of post-human warfare. On the other hand, it is a natural step in the evolution towards increasing the distance between combatants, and technology is only a prosthetic extinction of a human body (as Marshall McLuhan said). The paper raises the controversies over the creation of “computer morality”, focusing on the challenges and the possible ways to overcome them. Much attention is paid to the problem of laws of robotics, and the starting point is Asimov’s Laws.

### **Bibliografia:**

*A call for moratorium on the development and use of lethal autonomous robots*, United Nations Human Rights, Office of the High Commissioner, <http://www.ohchr.org/EN/NewsEvents/Pages/Acallformoratoriumonthedevelopmentrobots.aspx>, [25.02.2016].

Adams T.K., *Future Warfare and the Decline of Human Decisionmaking*, „Parameters”, 2001, nr 31:4, s. 57-71.

Asaro P.M., *How Just Could a Robot War Be?*, w: *Current Issues in Computing and Philosophy. Frontiers in Artificial Intelligence and Applications*, Briggie A., Waelbers K., Brey P. (red.), IOS Press, Amsterdam 2008.

Asimov I., *Robots and Empire*, <http://www.ebooktrove.com/Asimov,%20Isaac/Asimov,%20Isaac%20-%20Robot%2006%20-%20Robots%20&%20Empire.pdf>, [27.02.2016].

Asimov I., *Zabawa w berka*, w: Asimov I., *Ja, Robot*, Limbus, Bydgoszcz 1993.

Campa R., *Kodeksy etyczne robotów: zagadnienie kontroli sprawowanej przez człowieka*, „Pomiary Automatyka Robotyka”, 2001, nr 3.

*Campaign to Stop Killer Robots*, <https://www.stopkillerrobots.org/>, [23.02.2016].

Clarke R., *Asimov’s Laws of Robotics. Implications for Information Technology*, w: *Machine Ethics*, M. Anderson, S.L. Anderson (red.), Cambridge University Press, Cambridge 2011.

Crootof R., *The Killer Robots Are Here, Legal and Policy Implications*, „Cardozo Law Review”, 2015, nr 36:1837.

Danet D., Hanon J.-P., *Digitization and Robotization of the Battlefield: Evolution or Robolution?*, w: *Robots on the Battlefield, Contemporary Issues and*

*Implications for the Future*, red. Doaré R. i in., Combat Studies University Press, US Army Combined Arms Center, Fort Leavenworth 2014.

Erlanger S., *A Gaza War Full of Traps and Trickery*, „New York Times”, 2009, <http://www.nytimes.com/2009/01/11/world/middleeast/11hamas.html?partner=permalink&expod=permalink&r=0>, [28.02.2016].

Gips J., *Towards the ethical robot*, w: *Machine Ethics*, M. Anderson, S.L. Anderson (red.), Cambridge University Press, Cambridge 2011.

Griffits J., *Collateral damage: A brief history of U.S. mistakes at war*, CNN, <http://edition.cnn.com/2015/10/06/middleeast/us-collateral-damage-history/>, [27.02.2016].

Kamieński Ł., *Nowy wspaniały żołnierz. Rewolucja biotechnologiczna im wojna w XXI wieku*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2014.

Kamieński Ł., *Technologia i wojna przyszłości. Wokół nuklearnej i informacyjnej rewolucji w sprawach wojskowych*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2009.

Kopeć R., *Dyplomacja dronów*, „Kultura i Polityka”, 2015, nr 17.

Krishan A., *Killer Robots. Legality and Ethicality of Autonomous Weapons*, Ashgate, Farnham 2009.

McLuhan M., *Understanding Media: The Extension of Man*, McGraw-Hill Nowy Jork 1964.

Parker G., *Historia sztuki wojennej. Od starożytności do czasów współczesnych*, Książka i Wiedza, Warszawa 2008.

Scharre P., *Robotics on the Battlefield Part II: The Coming Swarm*, Center for a New American Security, Waszyngton 2014.

Scharre P., *Robotics on the Battlefield. Part I: Range, Persistence and Daring*, Center for a New American Security, Waszyngton 2014.

Sharkey N.E., *The inevitability of autonomous robot warfare*, „International Review of the Red Cross”, 2012, nr 94(886).

Windeck A., *Preface*, w: *Robots on the Battlefield, Contemporary Issues and Implications for the Future*, red. Doaré R. i in., Combat Studies University Press, US Army Combined Arms Center, Fort Leavenworth 2014.