

## Rafał Kopec\*

Uniwersytet Pedagogiczny w Krakowie

### ETYKA ROBOTÓW BOJOWYCH

Wykorzystanie robotów na polu walki stanowi jeden z najbardziej obiecujących trendów rozwoju techniki wojskowej, zwiastując rewolucję w sposobie prowadzenia wojen (robolecję). Roboty charakteryzują się wieloma przewagami w stosunku do ludzkich żołnierzy, które, co ciekawe, dotyczą nie tylko sfery skuteczności, ale potencjalnie także wymiaru etycznego. Zastosowanie maszyn będących robotami w pełnym znaczeniu tego słowa, czyli charakteryzujących się możliwością działania autonomicznego, zdolnych przy tym do podejmowania decyzji o selekcji i porażeniu celów (czyli decyzji dotyczących życia i śmierci), wymaga jednak rozwiązania szeregu dylematów sytuujących się na styku techniki i etyki. Artykuł koncentruje się na ukazaniu wyzwań oraz potencjalnych rozwiązań w zakresie budowy architektury sztucznej moralności dla autonomicznych systemów bojowych. Wysiłki w tym kierunku dzielą się na podejście odgórne i oddolne oraz stanowią przy tym odbicie dwóch stanowisk etycznych: rygorystycznego i sytuacyjnego.

Słowa kluczowe: robolecja, robot bojowy, autonomia, etyka robotów, sztuczna moralność

#### WPROWADZENIE

Historia wojen i wojskowości nie jest procesem linearnym. Podobnie jak w rozwoju nauki zdarzają się momenty zmiany paradygmatu (Khun 1968), tak na polu bitwy co pewien okres mamy do czynienia z rewolucyjnymi zmianami przekształcającymi charakter wojny, tzw. rewolucjami w sprawach wojskowych. Termin ten został wprowadzony przez Michaela Robertsa (1956) w kontekście rewolucyjnych zmian, jakie miały miejsce na europejskich polach bitew pomiędzy 1560 a 1660 rokiem. Rewolucje polegają na równoległej transformacji trzech elementów: techniki, doktryny (sposobu prowadzenia operacji wojskowych) oraz struktury organizacyjnej sił zbrojnych (Morningstar 1997: 37). Zazwyczaj jeden z tych elementów (często technika) jest dominujący, ewentualnie stanowi on „motor” przekształceń, pociągając za sobą zmiany w pozostałych sferach.

Być może stoimy obecnie u progu kolejnej rewolucji, związanej z zastosowaniem robotów w działaniach zbrojnych. Zwolennicy takiego podejścia mówią o tzw. robolecji – *robolecja* (Windeck 2014: V). Termin ten, powstały z połączenia słów robotyzacja i rewolucja, opiera się na założeniu, że robotyzacja prowadzi do rewolucyjnego przekształcenia charakteru wojny

---

\* Adres do korespondencji: Rafał Kopec, Instytut Bezpieczeństwa i Edukacji Obywatelskiej, Uniwersytet Pedagogiczny w Krakowie, ul. Ingardena 4, 30-060 Kraków; e-mail: rkopec@up.krakow.pl.

i stanowi kolejny punkt zwrotny w historii militarnej. Co więcej, zmiany w charakterze wojny powodowane przez roboty są bardziej dogłębne niż wcześniejsze rewolucje – przeobrażeniu ulega bowiem nie tyle sposób, w jaki prowadzi się wojnę, ile podmiot ją prowadzący – robot zamiast człowieka (Danet i Hanon 2014: XIII).

Roboty bojowe, mimo że znajdują się na początku drogi rozwojowej, już zdążyły zaważnąć zbiorową wyobraźnię. W szerszym ujęciu ich zastosowanie może być postrzegane jako ostatnia inkarnacja zachodniego sposobu prowadzenia wojen (*Western way of warfare*), charakteryzującego się technocentryzmem, czyli odwoływaniem się do przewagi technologicznej w celu osiągnięcia korzystnego rozstrzygnięcia na polu walki (Keegan 1993; Hanson 2001; Parker 2008; Kopeć 2014). Jest ono również jednym z przejawów charakterystycznego dla armii z zachodniego kręgu kulturowego dążenia do ograniczenia ryzyka dla własnych żołnierzy (*risk-transfer militarism*; Shaw 2002: 347), często będąc przy tym postrzegane jako typowa broń bogatych, ale pozbawionych heroizmu społeczeństw (w kontekście konfliktu mentalności heroicznej i postheroicznej Edwarda Luttwaka; 1995). Roboty bojowe stanowią odwrotność, a zarazem zaskakujące dopełnienie zamachowców samobójców, prezentując dwie skrajnie odmienne drogi do osiągnięcia tego samego celu – stworzenia ostatecznej broni precyzyjnej.

Robotyzacja pola walki – chociaż trudno wciąż określić, jaką dokładnie postać przybierze – budzi wiele obaw. Ich źródłem nie jest bynajmniej strach przed tym, że roboty będą miały możliwość zabijania. Tę zdolność posiadają już najprostsze systemy automatyczne, z minami na czele. Robotyzacja w pełnym znaczeniu tego słowa oznacza jednak proliferację maszyn, które będą miały zdolności utożsamiane z myśleniem. Autonomiczne roboty bojowe będą więc w stanie nie tylko zabijać, ale i decydować o życiu i śmierci.

Obawy te stanowią oś niniejszego artykułu. Wychodząc od definicji robota, ukazuje on powody, które mogą zdecydować o szczególnym kierunku robotyzacji, czyli autonomizacji (pod tym pojęciem kryje się budowa i zastosowanie autonomicznych systemów bojowych). Za tendencją tą przemawia nie tyle fakt, że dopiero system autonomiczny zasługuje w pełni na miano robota, ile rozliczne powody natury technicznej, warunkujące potencjalnie większą skuteczność na polu walki. Tendencja ta wiąże się jednak z dylematami sytuującymi się na styku techniki i etyki. Zasadniczym celem artykułu jest ukazanie tychże dylematów, a także wskazanie potencjalnych sposobów ich rozwiązania przez tworzenie architektury sztucznej moralności dla autonomicznych systemów bojowych.

## ROBOTYZACJA – JAK JĄ ROZUMIEĆ?

Słowo robot zostało po raz pierwszy zastosowane przez czeskiego pisarza Karela Čapka. Pojawiło się w sztuce *R.U.R. (Rossumovi Univerzální Roboti) – Uniwersalne roboty Rossuma* z roku 1921 (Čapek 2004). Roboty Čapka były jednak sztucznie produkowanymi istotami stworzonymi na podłożu biologicznym, czyli organizmami żywymi. Współczesna wizja robota zakładająca, że jest on istotą elektromechaniczną, pojawiła się w opowiadaniu Isaaca Asimova *Runaround (Zabawa w berka)* z 1942 roku (Asimov 2013). Robot jest więc maszyną (czyli urządzeniem technicznym wykorzystującym proces przekształcania energii), a dokładniej fuzją maszyny i komputera.

Według popularnej definicji robot to maszyna, która czuje, myśli i działa. Innymi słowy, robot jest komputerem (jego zadanie to myślenie) wyposażonym w sensory (odpowiedzialne za odczuwanie) i efektory (odpowiedzialne za działanie) (Lin, Bekey i Abney 2008: 101). Dzięki sensorom robot nie wymaga do funkcjonowania bezpośredniej interwencji człowieka (jak komputer wymagający np. operacji na klawiaturze czy myszce), lecz ma zdolność kierowania się bodźcami płynącymi bezpośrednio ze środowiska. Efektory z kolei warunkują zdolność do niecyfrowego oddziaływania na środowisko, w przeciwieństwie do komputera, który sam w sobie jest zdolny tylko do oddziaływania cyfrowego, czyli manipulacji danymi. Robot jest natomiast zdolny do manipulacji obiektami fizycznymi.

O ile określenie istoty odczuwania i działania nie stanowi dużego problemu, o tyle pojawia się on w przypadku definicji myślenia. Po pierwsze, odnosi się on do samego określenia, czym jest myślenie i czym jest inteligencja jako miara jego sprawności. Drugi problem kryje się w pytaniu: czy maszyny mogą myśleć, czy tylko są zdolne co najwyżej do symulowania myślenia (jak w eksperymencie myślowym Johna R. Searle'a, tzw. chińskim pokoju; Searle 1980: 417–424)? W celu zoperacjonalizowania pojęcia „robot” zamiast kategorii myślenia można zastosować kategorię autonomii, rozumianą w uproszczeniu jako zdolność do samodzielnego kształtowania swojego zachowania. Analizując to pojęcie bardziej dogłębnie, system autonomiczny winien charakteryzować się następującymi cechami: być niezależnym od operatora, mieć możliwość podejmowania decyzji w sposób swobodny oraz posiadać zdolność do funkcjonowania w nieprzewidywalnym środowisku (Marra i McNeil 2012: 1151–1152). Problemem pozostaje oczywiście stopniowanie tych cech i wyznaczenie takiego ich natężenia, które stanowiłoby próg pozwalający mówić o autonomii w przypadku danego systemu.

Robotyzację pola walki w sensie szerokim można rozumieć jako proliferację maszyn bezzałogowych (*unmanned* – pozbawionych załogi sterującej maszyną, zlokalizowanej na pokładzie; nie znaczy to, że maszyna musi być *uninhabited* – „niezamieszkana”, czyli bez człowieka na pokładzie, gdyż pojawiają się już maszyny bez załogi, ale zdolne do przewozu ludzi na przykład w celu ewakuacji rannych) (Scharre 2014: 10). Do urządzeń bezzałogowych zaliczymy między innymi maszyny zdalnie sterowane (na przykład roboty saperskie, będące jedną z najpopularniejszych kategorii robotów na współczesnym polu walki) czy też broń kierowaną (określaną często, zdecydowanie na wyrost, inteligentną; Pucket 2004: 648). Tak rozumiana robotyzacja jest jednak tylko kolejnym etapem zwiększenia dystansu między walczącymi. Prawdziwa robolucja będzie się wiązać z zastosowaniem robotów autonomicznych, a w szczególności autonomicznych robotów bojowych, zdolnych do samodzielnej selekcji celów oraz ich zwalczania bez potrzeby „pytania o zdanie” operatora, a nawet pozbawionych możliwości odwołania się do jego opinii.

## DLACZEGO AUTONOMIA?

Jakie zalety mają systemy autonomiczne? Spełnienie wymagań autonomii jest wszakże ogromnym wyzwaniem technicznym, a tymczasem już proste systemy zdalnie sterowane pozwalają osiągnąć zasadniczy cel robotyzacji – ograniczyć ryzyko dla żołnierzy, przez eliminację człowieka z pokładu maszyny podczas wykonywania najbardziej niebezpiecznych zadań.

Najpoważniejszą wadą zdalnego sterowania jest możliwość zakłócenia kanałów łączności albo wręcz przejęcia kontroli nad nimi przez przeciwnika (Kim, Wampler, Goppert et al. 2015). Sprowadzenie na ziemię i przejęcie w stanie praktycznie nieuszkodzonym amerykańskiego drona rozpoznawczego RQ-170 Sentinel przez Iran (Hartman i Steup 2013: 7–8) dobitnie pokazuje, że jest to niebezpieczeństwo realne. W obliczu przeciwnika dysponującego określonymi możliwościami w zakresie walki elektronicznej systemy zdalnie sterowane mogą okazać się wręcz bezużyteczne, nie pozwalając na realizację wyznaczonych im zadań.

Znaczącą wadą, zwłaszcza w przypadku stosowania łącz satelitarnych, jest opóźnienie w przesyłaniu komend kierowania. Opóźnienie wynoszące około sekundy, chociaż wydaje się marginalne, to stawia pod znakiem zapytania zdolność do wykonywania niektórych zadań. Trudno sobie na przykład wyobrazić samolot zdalnie sterowany zdolny do toczenia równorzędnej walki powietrznej z samolotem pilotowanym. Oprócz opóźnienia, w grę wchodzi tu czynnik izolacji od maszyny i warunków środowiskowych, co utrudnia prawidłowe „czucie” maszyny (Wallace 2009). Z tych powodów układem optymalnym wydaje się umieszczenie jednostki podejmującej decyzję (operatora czy pilota w przypadku maszyn załogowych, bądź też systemu tzw. sztucznej inteligencji w przypadku maszyn bezzałogowych) na pokładzie maszyny.

Jeśli przeanalizować szczegółowo kryteria autonomii, okazuje się, że spełniające je systemy bojowe już istnieją. Chodzi przede wszystkim o zaawansowane systemy przeciwlotnicze i przeciwrakietowe, a także o tzw. robotów-strażników. Są to jednak jak dotychczas systemy defensywne i stacjonarne (w tym sensie, że niezdolne do autonomicznego poruszania się – niektóre systemy, na przykład okrętowe, są bowiem zamontowane na mobilnych platformach, ale mobilność nie jest w ich przypadku funkcją poddaną autonomizacji). Dodatkowo kryterium zdolności do funkcjonowania w niepewnym środowisku jest spełnione w sposób dyskusyjny – nie ma bowiem jednoznacznie określonego progu, który wyznaczałby odpowiedni dla systemów autonomicznych stopień komplikacji i nieprzewidywalności otoczenia. Prawdziwa rewolucja robotyczna rozpocznie się jednak z chwilą, gdy pojawią się systemy mobilne i ofensywne, spełniające przy tym wszystkie wymienione kryteria autonomii. Zastosowanie takich systemów wiązało się będzie z koniecznością pokonania problemów technicznych oraz rozstrzygnięcia dylematów etycznych.

## W CZYM ROBOTY SĄ LEPSZE OD LUDZI?

Fakt, że roboty (czy też komputery, jeśli koncentrujemy się na elemencie odpowiedzialnym za „myślenie”) radzą sobie z pewnymi zadaniami lepiej od ludzi, nie ulega wątpliwości. Dotyczy to operacji „umysłowych” – wszak przetwarzanie informacji jest tym, w czym komputery są najlepsze. Przykładem może być umiejętność gry w szachy – pierwsze programy komputerowe służące do tego celu powstały już w latach 50. ubiegłego wieku (Bernstein i Roberts 1958), a wydarzeniem o znaczeniu symbolicznym było zwycięstwo komputera Deep Blue z mistrzem świata Garrim Kasparowem w 1997 roku (Pandolfini 1997). Roboty przewyższają też ludzi pod względem fizycznym – są w stanie operować w środowiskach niedostępnych dla człowieka, nie odczuwają zmęczenia, często cechują się lepszymi parametrami w zakresie na przykład siły czy szybkości.

Roboty potencjalnie mogą być również lepsze od żołnierzy nie tylko pod względem skuteczności, ale także z uwagi na uwarunkowania etyczne. Czynnikiem stanowiącymi potencjalnie silne strony robotów w kontekście stosowania zasad etycznych na polu walki są (Arkin 2010: 333):

- brak konieczności chronienia samego siebie (co może oznaczać nawet samopoświęcenie), co pozwala zwiększyć zdolność właściwej identyfikacji, na przykład w sytuacji konieczności zbliżenia się do nierozpoznanego obiektu; roboty nie muszą działać na zasadzie „najpierw strzelaj, potem pytaj”;
- brak czynników emocjonalnych mogących zaburzać osąd (strach, histeria, panika itp.);
- uniknięcie psychologicznego syndromu samospelniającej się przepowiedni; w sytuacji stresowej ludzie mają tendencję do akceptacji informacji, które pasują do przyjętego uprzednio wzorca postępowania i negocjowania tych, które są z nim sprzeczne;
- duże możliwości instalacji sensorów (optoelektronika, radary, czujniki sejsmiczne);
- zdolność do szybszego przetwarzania informacji z wielu źródeł w jednostce czasu; w wielu przypadkach oznacza to, że robot zdąży przeanalizować informacje, zanim podejmie decyzję o użyciu broni;
- możliwość monitorowania pod względem etycznym działań osób współpracujących z robotem.

W efekcie „niehumanizacji wojny”, potencjalnie przynajmniej, mogą funkcjonować na polu walki w sposób bardziej ludzki niż ludzie. Co więcej, autonomizacja rozumiana jako cedowanie na robota decyzji o otwarciu ognia (decyzji w zakresie wyboru celów i ich rażenia) może doprowadzić do odwrócenia tendencji charakterystycznej dla procesu rozwoju platform bezzałogowych, sterowanych przez człowieka. Rezultatem tego procesu jest zwiększenie dystansu walki. Jest to warunkowane dążeniem do zmniejszenia ryzyka dla żołnierzy, ale powoduje też „izolację” od procesu zabijania. Zwiększenie dystansu fizycznego pociąga więc za sobą zwiększenie dystansu psychologicznego (Grossman 2010: 185–189). Żołnierz przestaje odczuwać, że efektem jego czynów jest śmierć innych osób, często nie widzi nawet konsekwencji swego działania, prowadzi tzw. wojnę z wygodnego fotela (*armchair warfare*; Olson i Rashid 2013: 2). Przykładem tego zjawiska jest tzw. mentalność *play station* (Simons 2011: 11), o prezentowanie której oskarża się operatorów uzbrojonych dronów. Zakłada ona, że sterowanie dronem z bezpiecznej bazy w Stanach Zjednoczonych powoduje postrzeganie rzeczywistości na podobieństwo gry komputerowej i idącą za tym nadmierną tendencję do szybkiego i często bezrefleksyjnego otwierania ognia.

Roboty mogą stanowić rozwiązanie dwóch kluczowych problemów związanych z uczestnictwem ludzi w działaniach wojennych. Pierwszym jest możliwość uniknięcia szkód psychicznych, w tym syndromu stresu pourazowego, będących rezultatem uczestnictwa człowieka w działaniach na polu walki, a w szczególności wystawienia na okoliczności związane z zabijaniem (Krishan 2009: 124). Po drugie, roboty mogą stać się sposobem na przezwycięzenie występującego często wśród żołnierzy braku ducha ofensywnego spowodowanego względami psychologicznymi (nie chodzi tylko o strach, ale również o awersję do zabijania – część żołnierzy nie jest zdolna strzelać do przeciwnika (Stout 2005: 66); co ciekawe, zjawisko to dotyczy nawet pilotów samolotów myśliwskich (Arkin 2010: 339).

## POZA KONTROLĄ – NIEBEZPIECZEŃSTWA AUTONOMIZACJI

Potencjalne problemy etyczne związane z autonomicznymi systemami bojowymi można podzielić na następujące kategorie (Kuptel i Williams 2014: 19–22).

1. Niewłaściwe działanie systemu, czyli działanie niezgodne z intencją podmiotu podejmującego decyzję o zastosowaniu systemu. Chodzi zarówno o „oczywiste” awarie (jak w przypadku awarii automatycznie kontrolowanej armaty przeciwlotniczej 35 mm w Republice Południowej Afryki w 2007 roku, która na skutek błędu systemu sterowania otwarła ogień, zabijając dziewięciu żołnierzy w czasie ćwiczeń (Meadows 2011: 26). Zasadniczym problemem jest jednak możliwość podjęcia błędnych decyzji dotyczących selekcji i rażenia celów, które mogą prowadzić do strat wśród ludności cywilnej lub własnych żołnierzy.

Największym problemem jest rozróżnienie między kombatantami i niekombatantami (walczącymi i niewalczącymi). Prawo wojny zakłada, że uprawnione jest tylko intencjonalne atakowanie walczących, czyli osób posiadających status kombatanta (żołnierza) oraz cywilów biorących bezpośredni udział w walkach (Murphy i Radsan 2009: 409). We współczesnych konfliktach zbrojnych noszenie munduru (jeden z atrybutów żołnierza) coraz częściej zamiast reguły staje się wyjątkiem (Foreman 2013: 930). Sytuacja ta jest pochodną niejednoznacznego statusu wielu współczesnych konfliktów zbrojnych – podczas gdy prawo międzynarodowe rozróżnia tylko konflikty międzynarodowe (międzypaństwowe) oraz wewnętrzne (w ramach jednego państwa) (Daskal 2013: 1175), wiele współczesnych konfliktów nie mieści się w ramach tej dychotomii i zaliczyć je można do nieobjętej regułami prawa międzynarodowego kategorii konfliktów wewnętrznych umiędzynarodowionych (*extrastate*) (Schöndorf 2004: 3–4). W rezultacie nie wiadomo dokładnie, jakie reguły prawne są adekwatne w ramach tego typu konfliktu. Co za tym idzie, nie sposób jednoznacznie określić, kto powinien być traktowany jako osoba niewalcząca, a komu przysługuje status osoby walczącej. Status taki wiąże się zarówno z uprawnieniami, na przykład do praw jeńca w razie schwytania, jak i z konsekwencjami w postaci nieustannego narażenia na atak – nie tylko w sytuacji zaangażowania w bezpośrednie działania. W przypadku robotów bojowych problem ten jest pochodną zarówno niedoskonałości sensorów, jednak w większym stopniu jest to rezultat trudności w napisaniu programu komputerowego, który radziłby sobie z takim zadaniem. W takich sytuacjach potrzebna jest świadomość sytuacyjna i pewna doza zdrowego rozsądku, której brakuje maszynom. Stąd pojawiające się propozycje zakazu robotów swobodnie poruszających się, zdolnych do autonomicznego wyboru celów i letalnego oddziaływania na cel (*autonomous lethal targeting by free-ranging robots*) (Sharkey 2012: 788–789). Z uwagi na uwzględnienie w działaniach z użyciem robotów bojowych zapisów międzynarodowego prawa humanitarnego pojawia się również kwestia nakazu proporcjonalności użytych środków, w tym zakazu powodowania niepotrzebnych cierpień (Kanwar 2011: 620). W tym przypadku jednak z pomocą przyjsć mogą stosowane już w siłach zbrojnych algorytmy szacujące możliwe skutki uboczne wskutek rozważanego ataku raketowego lub za pomocą dronów (Crawford 2013: 351). Analogiczny algorytm może być opracowany dla robotów autonomicznych (Gubrud 2014: 35).

2. Niewłaściwe użycie systemu, czyli intencjonalne użycie w sposób odmienny od tego, w jaki system został zaprojektowany, na przykład użycie w nielegalnych akcjach, do realizacji działań stanowiących zbrodnię wojenną. Taka możliwość dotyczy w zasadzie każdej broni, nawet najbardziej prymitywnej, ale w przypadku robotów pojawia się dodatkowy dylemat – dylemat rozmytej odpowiedzialności. Zasadnicze pytanie brzmi: czy autonomiczne roboty bojowe mogą być potraktowane jako żołnierze (skoro same podejmują decyzje, powinny też za nie odpowiadać), czy też zawsze należy traktować je jako broń? W przypadku tej drugiej odpowiedzi nie sposób nie zadać kolejnego pytania – kto w takiej sytuacji ponosi odpowiedzialność? Czy zawsze będzie to osoba, która delegowała robota do wykonania danego zadania?

Wydaje się, że ewentualne traktowanie robotów jako żołnierzy jest jednym z przejawów ich nadmiernej antropomorfizacji. Systemy autonomiczne powinny być postrzegane jako „agenci kontrolowani przez konkretnych ludzi”. Słowo odpowiedzialność zakłada bowiem występowanie działania dobrowolnego i świadomie intencjonalnego. Dotyczy więc tylko ludzi, bo tylko oni mogą działać w sposób wolny i być zdolni do przewidywania konsekwencji (Johnson i Noorman 2014: 18). Delegowanie zadań dla systemów autonomicznych powinno być równoznaczne z braniem odpowiedzialności za ich działanie.

3. Nieoczekiwane konsekwencje. Proliferacja autonomicznych systemów bojowych może pociągnąć za sobą następujące konsekwencje w szerszym polityczno-społecznym wymiarze:

- zmniejszenie bariery przed podjęciem decyzji o wojnie, wynikające z obniżenia psychologicznego progu wstrzymującego przed rozpoczęciem działań bojowych, w sytuacji gdy nie ryzykuje się życia własnych żołnierzy (tzw. *risk-free warfare*) (Asaro 2008: 56–58);
- zachęta do atakowania cywilów przez stronę, która nie może atakować żołnierzy, bo zastąpieni zostali robotami;
- możliwość deprecjacji tradycyjnych profesjonalnych umiejętności żołnierza; w sytuacji upowszechnienia robotów umiejętności te jawić się będą jako niepotrzebne i anachroniczne;
- możliwość rozwoju sposobów walki określanych jako *lawfare*, polegających na wykorzystaniu prawa jako narzędzia na polu walki, na przykład uzyskaniu taktycznej przewagi dzięki znajomości reguł otwarcia ognia stosowanych przez przeciwnika (Singer 2010: 402); przeciwnik, znając reguły etycznego działania, wedle których działa robot, może wykorzystać tę wiedzę, na przykład tak aranżując sytuację na polu walki, by robot był „zmuszony” do powstrzymania się od działania (Magnuson 2009: 29);
- trudność w wygraniu „wojny o serca i umysły” z wykorzystaniem metod uznawanych przez wiele społeczeństw za niehonorowe (Kaplan 2013: 41–42);
- potencjalny negatywny wpływ na spójność jednostek bojowych; obecność pośród żołnierzy robotów wyposażonych w sensory i mogących służyć potencjalnie do monitorowania działań ludzi, może podważać zaufanie w ramach zespołu (Lin, Abney i Bekey 2013: 356).

## ARCHITEKTURA PĘTLI DECYZYJNEJ

Jeśli robot działa tak, jak go zaprogramowano, można określić go jako „operacyjnie moralny”. Takie spojrzenie może być jednak niewystarczające w sytuacji, gdy roboty stają się coraz bardziej autonomiczne, a środowisko, w którym funkcjonują, może być coraz bardziej złożone i kompleksowe (w sensie różnorodności bodźców). Trudno więc dokładnie przewidzieć reakcje robota na sygnały płynące ze środowiska. Potrzebna jest więc moralność funkcjonalna, czyli zdolność do brania pod uwagę uwarunkowań moralnych. Powstaje jednak pytanie, jak te moralne uwarunkowania mają być określane, skoro etyka jest uzależniona od kontekstu kulturowego. Roboty powstające w różnych kulturach będą cechowały się więc odmienną moralnością funkcjonalną. Rozwiązaniem może być związanie tego typu moralności z regułami prawa międzynarodowego.

W tym momencie pojawia się jednak dylemat, w jaki sposób abstrakcyjne reguły moralne zaimplementować w inteligentny system (dylemat dotyczy też człowieka – tego, w jaki sposób staje się on istotą moralną). Generalnie wyróżnić można dwa podejścia: odgórne (*top-down*) i oddolne (*bottom-up*).

Podejście odgórne oznacza zakodowanie danej teorii etycznej w schemat działania (w przypadku robotów – w oprogramowanie). Z kolei podejście oddolne polega na rozwoju moralności w wyniku nabywania doświadczeń (w uproszczeniu chodzi o podobny proces, jak ten, w ramach którego dzieci stają się dorosłymi) (Lin, Bekey i Abney 2008: 27). Podejścia te można porównać do dwóch modeli szkolenia żołnierzy w zakresie reguł otwarcia ognia (*rules of engagement*, ROE) – tzw. *legislative model*, oparty na wdrożeniu reguł zewnętrznych w wyniku analizy i interpretacji skodyfikowanych (spisanych) reguł, oraz tzw. *training model*, polegający na wykorzystaniu reguł wewnętrznych, już wcześniej zinternalizowanych przez szkoloną osobę, i ich praktycznej implementacji w postaci odpowiedniego zachowania w danej sytuacji (tzw. *behavioral performance*) (Arkin 2007: 68).

Podejście odgórne zakłada dekompozycję zadania w prostsze subzadania. Komponenty moralności są grupowane w moduły odpowiedzialne za wykonanie subzadań i następnie hierarchicznie układane, by realizować zadania wyższego rzędu. W podejściu *bottom-up* z kolei nacisk jest położony na wykreowanie środowiska, gdzie maszyna stosowałaby różne schematy działania, następnie dokonywała ewaluacji, wybierając działanie, które byłoby „moralnie opłacalne”. Reguły moralne byłyby więc odkrywane i konstruowane przez samą maszynę (Lin, Bekey i Abney 2008: 28). Ta dychotomia jest oczywiście w pewnym stopniu upraszczająca, jednak dylemat dotyczący efektywności tych podejść w budowaniu sztucznej moralności jest jednym z centralnych zagadnień związanych z rozwojem autonomicznych systemów bojowych.

Podejście odgórne zakłada oparcie się na zespole reguł, które mogą być przetransformowane w algorytm. Podejście to jest odzwierciedleniem deontologii w etyce (nauce o powinnościach i obowiązkach; kantyzm jest rodzajem deontologii, a sam Kant ojcem tego nurtu filozofii moralnej, Ślipko 2004: 83–85). Tego rodzaju etyka zakłada, że pewnych czynów absolutnie nie należy dokonywać (cel nie uświęca środków), nawet jeśli potencjalnie prowadziłyby to do pozytywnych konsekwencji.

Zespół reguł będący podstawą tego rodzaju etycznej autonomii powinien charakteryzować się następującymi cechami (Arkin 2007: 39).



- Niesprzeczność: poszczególne reguły będące podstawą dyspozycji do działania nie powinny znajdować się ze sobą w konflikcie. Przykładem podejścia odgórnego są prawa robotyki Asimova (najbardziej znany literacki przykład rozwiązania problemu etyki robotów). Poszczególne reguły są hierarchizowane w taki sposób, by uniknąć konfliktu między nimi, ale jak pokazują same opowiadania Asimova, nie w każdym przypadku będzie to możliwe (Lin, Bekey i Abney 2008: 30).
- Kompletność: reguły powinny zawierać wskazówki co do działania w przypadku każdego dylematu etycznego.
- Praktyczność: dyspozycje do działania wynikające z reguł muszą być możliwe do zastosowania.

Działania będące rezultatem dyspozycji podzielić można na trzy grupy:

- obligatoryjne: agent jest z powodów etycznych zobowiązany do podjęcia akcji;
- dopuszczalne: działanie jest moralnie dopuszczalne, ale nie jest wymagane;
- zabronione: działanie jest moralnie niedopuszczalne.

Między tymi działaniami zachodzą następujące relacje logiczne (Arkin 2007: 40):

- jeśli działanie jest dopuszczalne, jest potencjalnie obligatoryjne, ale nie jest zabronione;
- jeśli działanie jest obligatoryjne, jest dopuszczalne i nie jest zabronione;
- jeśli działanie jest zabronione, nie jest dopuszczalne i nie jest obligatoryjne.

Według Rona Arkina, działając moralnie, agent nigdy nie powinien podejmować zabronionej akcji śmiertecznej. Co więcej, sama dopuszczalność działania nie jest wystarczająca. Robot może podjąć działanie wyłącznie wtedy, gdy jest ono również obligatoryjne w kontekście militarnej konieczności. Arkin zakłada więc, że działania śmierteczne systemów autonomicznych powinny być zawsze obligatoryjne, a nie tylko dopuszczalne. Algorytm wyznaczający funkcjonowanie robotów bojowych powinien być przy tym oparty na dwóch zespołach reguł moralnych. Pierwszym z nich są prawa wojny (*law of war*), które opisują w sposób ogólny absolutnie zabronione działania śmierteczne, wskazując to, co jest jednoznacznie zakazane. Drugim komponentem są reguły otwarcia ognia (*rules of engagement*) konkretnej operacji, które definiują specyficzne dla sytuacji zasady w zakresie podejmowania działania śmiertecznego. Tym samym wskazują, czy dane działanie w kontekście celów operacji można uznać za obligatoryjne, co stanowiłoby zobowiązanie do jego podjęcia (Arkin 2007: 40).

Podejście oddolne opiera się z kolei na dwóch zasadach: złożeniu kompletnych systemów (umiejętności) z odrębnych subsystemów oraz pojawianiu się reguł i wzorów zachowań wraz z procesem eksperymentowania, które podejmuje sztuczny byt. Proces ten ma doprowadzić do agregacji odrębnych umiejętności w umiejętności wyższego rzędu o charakterze kognitywnym. To podejście czerpie inspirację z modelu ewolucji i zakłada samoselekcję i samoorganizację systemu w kierunku optymalizacji osiągnięć poprzez maksymalizację korzyści. W sensie konstrukcyjnym podstawą takiego funkcjonowania ma być oparcie systemu sztucznej inteligencji na sieciach neuronowych. Dostrzec można tu analogię do psychologii ewolucyjnej zakładającej, że takie cechy jak zdolność do kooperacji pojawiają się w toku ewolucji i mogą być reprodukowane w genach (Lin, Bekey i Abney 2008: 36).

Podjęcie oddolne zakłada pojawienie się dynamicznej moralności, będącej w stanie reagować w różnych sytuacjach. Stanowi ono odzwierciedlenie partykularyzmu – podejścia etycznego opierającego się na zasadzie braku ogólnej, uniwersalnej moralności, podkreślającego odmienność, unikalność i niepowtarzalność każdej sytuacji. Jest to rodzaj etyki sytuacyjnej, oceniającej postępowanie w kontekście danej sytuacji, w przeciwieństwie do etyki rygorystycznej, głoszącej istnienie absolutnych zasad moralnych. W przypadku robotów bojowych zasadniczy problem związany z podejściem oddolnym jest oczywisty: jeśli robot uczyć się ma moralności jak dziecko, istnieje znaczne prawdopodobieństwo, że na polu walki zachowa się w sposób niewłaściwy.

**Tabela 1.** Architektura etyczna autonomicznych systemów bojowych

<b>Podejście odgórne</b>	<b>Podejście oddolne</b>
Działanie według określonych z góry zasad	Samouczenie się zasad postępowania
Deontologia (zasady uniwersalne)	Partykularyzm (zasady partykularne)
Etyka rygorystyczna	Etyka sytuacyjna
Algorytm (zdefiniowany ciąg poleceń)	Sieci neuronowe
Robot jako narzędzie	Robot jako agent

Źródło: opracowanie własne

## W KIERUNKU SZTUCZNEJ INTUICJI MORALNEJ

Roboty, nawet jeśli uda się zaimplementować w ich system sterowanie regułami moralnymi, pozbawione są emocji. Mają więc problem z interpretowaniem ludzkiego działania. W warunkach współczesnych nieliniarnych, asymetrycznych konfliktów może to prowadzić do deficytu świadomości sytuacyjnej, a w szczególności do problemów z rozróżnianiem osób walczących od niewalczących w nieoczywistych, wątpliwych sytuacjach (Gubrud 2014: 36). Do zrozumienia ludzkich intencji i zachowań potrzebna jest bowiem intuicja oraz empatia, a także zdolność wyciągania wniosków także na podstawie takich ulotnych przesłanek (Sharkey 2010: 279).

W przypadku ludzi decyzje dotyczące zagadnień moralnych oparte są na dwóch podłożach. Pierwszym jest intuicja moralna – działa szybko, automatycznie, bezwolnie, nie wymagając zbyt wiele uwagi, przez co jest odporna na przeciwstawne rozumowanie. Drugie podłoże to rozumowanie w zakresie zgodności z zasadami – działające wolniej, wymagające uwagi, dopuszczające rozmyślanie na tematy moralne, w pewnym stopniu otwarte na twierdzenia przeciwstawne, a przynajmniej zakładające możliwość rozważenia argumentów niezgodnych, a nawet opozycyjnych w stosunku do przyjętego systemu moralnego. Sądy moralne mogą być wydawane na podstawie jednego lub obydwu z tych komponentów (Arkin 2007: 55).

Według Arkina, by roboty były w pełni etyczne, należy „zaszczepić” im pewne sztuczne uczucia. Będą one stanowiły podłoże sztucznej intuicji moralnej. Przykładowo, jeśli robot

zachowałby się nieetycznie (na przykład wskutek wykonywania rozkazów ludzkich), aktywowałoby się w ramach jego systemu sterowania sztuczne poczucie winy (Arkin 2009: 66). W sytuacji gdyby pewien ustalony wcześniej próg natężenia tego sztucznego uczucia został przekroczony, robot mógłby odmówić użycia uzbrojenia.

## BŁĄD ANTROPOMORFIZACJI. CZY MOŻLIWY JEST ROBOT KANTOWSKI?

Sztuczne poczucie winy i schemat jego działania na pozór wydają się odwzorowaniem ludzkiego poczucia winy i schematu działania znanego ze *Zbrodni i kary* Fiodora Dostojewskiego. Sugeruje to, że robot może być istotą świadomą i może przeżywać stany emocjonalne. Tak naprawdę jednak bardziej zasadna jest analogia do termostatu – po przekroczeniu pewnego, z góry określonego progu (temperaturowego w przypadku termostatu, „uczuciowego” w przypadku robota) aktywowane bądź dezaktywowane byłoby określone działanie (Sharkey 2012: 795).

W przypadku robotów często mamy do czynienia ze swoistym antropomorfizmem (przypisywaniem ludzkich cech istotom, którym się one nie należą), wzmacnianym przez twórczość z gatunku *science fiction*. Opiera się on na założeniu, że roboty myślą i działają w sposób podobny do ludzi, ale zazwyczaj lepiej, bardziej sprawnie. Ten mit jest stosunkowo powszechny, zniekształcając przy okazji polityczne bądź prawne dyskusje o robotach.

Tymczasem roboty mogą co najwyżej charakteryzować się tzw. moralnością niewolnika – nie realizują własnych celów, lecz cele wyznaczone przez podmiot zewnętrzny (w przypadku robotów bojowych jest to osoba wydająca rozkazy) (Lin, Bekey i Abney 2008: 64). Roboty nie mogą stanowić swoich praw i reguł. Ich wolność dotyczy tylko wyboru środków, a nie celów.

Nawet jeśli roboty mają zdolności decyzyjne, nie mogą charakteryzować się autonomią w sensie kantowskim. Co za tym idzie, nie mogą brać odpowiedzialności za swoje czyny. Nie mogą być więc na przykład oskarżone o zbrodnie wojenne (pomijając oczywisty dylemat – w jaki sposób ukarać robota?). Moralność niewolnika nie jest przy tym utożsamiana z sytuacją niewolnika. Człowiek, nawet niewolnik albo żołnierz zmuszony do wykonywania rozkazów, zawsze ma wybór. W sytuacji dylematu etycznego człowiek ma możliwość sprzeciwienia się rozkazom (co może pociągnąć za sobą fatalne, nawet śmiertelne konsekwencje). Robot tymczasem jest zupełnie pozbawiony tej możliwości. Na tym polega moralność niewolnika – na braku możliwości działania sprzecznego z dyspozycjami. W przypadku robota niewłaściwe działanie jest zawsze pochodną niewłaściwego zaprogramowania.

Moralność niewolnika charakteryzuje nie tylko roboty funkcjonujące na podstawie architektury *top-down*, ale także roboty samouczące się. One także nie mają autonomii w sensie kantowskim, czyli możliwości wyboru „drogi życiowej” oraz celów do realizacji. Mogą jedynie wybierać środki do realizacji celów wyznaczonych przez kogoś innego. Być może dzięki zdolności samouczenia się będą mogły robić to lepiej, ale nie oznacza to autonomii w sensie kantowskim.

Nie znaczy to jednak, że zbudowanie robota autonomicznego w sensie kantowskim (*kantian-autonomous robot*) jest zupełnie niemożliwe. Tego rodzaju sztuczna inteligencja musiałaby jednak posiadać możliwość stworzenia własnego programu, który obejmowałby

także fundamentalne „cele życiowe” robota. Taki robot mógłby się wyzwolić z moralności niewolnika i być kreatorem swojego losu. Nie wiadomo, czy stworzenie takiego robota jest technicznie wykonalne – na pewno byłoby to niebezpieczne, a także niewskazane z militarnego punktu widzenia.

## WNIOSKI

Prawdziwa rewolucja związana z robotyzacją pola walki dokona się z chwilą upowszechnienia się systemów autonomicznych. Stosowanie systemów bezzałogowych, lecz nieautonomicznych (na przykład zdalnie sterowanych) jest przejawem ewolucji – tylko (i aż) kolejnym krokiem na drodze do zwiększenia dystansu pola walki. Zamyśl stojący za tym dążeniem jest jasny: zwiększyć bezpieczeństwo żołnierzy, w jak największym stopniu zminimalizować ryzyko wiążące się z uczestnictwem w działaniach bojowych. Stąd szerokie wykorzystanie systemów bezzałogowych (określanych powszechnie, lecz nie do końca poprawnie jako roboty) w realizacji zadań charakteryzujących się największym stopniem ryzyka. Przykładami mogą być działania rozpoznawcze (w tym w przestrzeni powietrznej kontrolowanej przez przeciwnika), a także działania saperskie (w tym rozpoznanie i likwidacja improwizowanych ładunków wybuchowych).

Robotyzacja w pełnym znaczeniu tego słowa, rozumiana jako szerokie wykorzystanie autonomicznych systemów bojowych, może jednak przynieść korzyści daleko wykraczające poza wyżej wymienione. Co ciekawe, mogą one sytuować się nie tylko w sferze militarnej użyteczności, lecz potencjalnie także w sferze etyki prowadzenia działań wojennych. Eksperci wskazują, że pewne cechy charakterystyczne dla robotów, i odróżniające ich od ludzkich żołnierzy, pozwolą im funkcjonować na polu walki w sposób często bardziej etyczny niż w przypadku ludzi.

By ta, być może nadmiernie optymistyczna wizja, ziściła się, potrzeba jednak rozwiązania szeregu problemów związanych z implementacją zasad etycznych w systemy sterowania robotów bojowych. Dwa zasadnicze kierunki prac nad systemami etycznymi robotów (podejście odgórne i oddolne) odpowiadają różnym sposobom budowania systemów etycznych przez człowieka. Różnymi drogami próbują one doprowadzić do stworzenia systemów umożliwiających adekwatne zachowanie robota (zarówno z punktu widzenia etyki, jak i realizacji zadania) w nieprzewidywalnym środowisku.

Systemy te jednak mogą okazać się niewystarczające z powodu braku czynnika emocjonalnego, który w interpretowaniu zachowania ludzi może być kluczowy. Pomysł „zaszczepienia” robotom sztucznych odczuć prezentowany przez Ronalda Arkina jest jednak kontrowersyjny i może być traktowany jako przejaw nieuprawnionej antropomorfizacji robotów. Zdaniem autora jest ona nieuprawniona, gdyż po pierwsze, obecne możliwości techniczne pozwalają jedynie na prymitywną symulację odczuć (Noel Sharkey porównuje ją pod względem zasady działania do termostatu), po drugie zaś, prowadzić może do przypisywania robotom właściwości charakterystycznych jedynie dla ludzi. Chodzi przede wszystkim o autonomię moralną i wszystko, co pociąga ona za sobą, ze zdolnością do dokonywania wyborów moralnych i gotowością do ponoszenia odpowiedzialności za swoje czyny na czele.

## BIBLIOGRAFIA

- Arkin, Ronald C. 2007. *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*, Atlanta: Georgia Institute of Technology.
- Arkin, Ronald C. 2009. *Governing Lethal Behavior in Autonomous Robots*, Boca Raton: A Chapman & Hall Book.
- Arkin, Ronald C. 2010. *The Case for Ethical Autonomy in Unmanned Systems*, „Journal of Military Ethics”, 9(4): 332–341.
- Asaro, Peter M. 2008. *How Just Could a Robot War Be?*, w: Adam Briggie, Katinka Waelbers i Philip Brey (red.), *Current Issues in Computing and Philosophy. Frontiers in Artificial Intelligence and Applications*, Amsterdam: IOS Press, s. 50–64.
- Asimov, Isaac. 2013. *Ja, robot*, Poznań: Rebis.
- Bernstein, Alex i Michael de V. Roberts. 1958. *Computer vs. Chess-Player*, „Scientific American”, June: 1–6.
- Čapek, Karel. 2004. *R.U.R. (Rossum's Universal Robots)*, New York: Penguin Classics.
- Crawford, Neta C. 2013. *Accountability for Killing. Moral Responsibility for Collateral Damage in America's Post-9/11 Wars*, Oxford: Oxford University Press.
- Danet, Didier i Jean-Paul Hanon. 2014. *Digitization and Robotization of the Battlefield: Evolution or Robolution?*, w: Ronan Doaré, Didier Danet, Jean-Paul Hanon i Gérard de Boisboissel (red.), *Robots on the Battlefield, Contemporary Issues and Implications for the Future*, Fort Leavenworth: Combat Studies Institute Press, s. XIII–XXXII.
- Daskal, Jennifer. 2013. *The Geography of the Battlefield: A Framework for Detention and Targeting Outside the 'Hot' Conflict Zone*, „University of Pennsylvania Law Review”, 161(5): 1165–1234.
- Foreman, Melanie J. 2013. *When Targeted Killing Is Not Permissible: An Evaluation of Targeted Killing Under the Laws of War and Morality*, „Journal of Constitutional Law”, 15(3): 921–960.
- Grossman, Dave. 2010. *O zabijaniu. Psychologiczny koszt kształtowania gotowości do zabijania w czasach wojny i pokoju*, Warszawa: Mayfly.
- Gubrud, Mark. 2014. *Stopping killer robots*, „Bulletin of Atomic Scientists”, 70(1): 32–42.
- Hanson, Victor Davis. 2001. *Carnage and Culture. Landmark Battles in the Rise to Western Power*, New York: Anchor Books.
- Hartman, Kim i Christoph Steup. 2013. *The Vulnerability of UAVs to Cyber Attacks – An Approach to the Risk Assessment*, NATO CCD COE Publications, [https://ccdcoe.org/cycon/2013/proceedings/d3r2s2\\_hartmann.pdf](https://ccdcoe.org/cycon/2013/proceedings/d3r2s2_hartmann.pdf) [18.02.2016].
- Johnson, Deborah G. i Merel E. Noorman. 2014. *Responsibility Practices in Robotic Warfare*, „Military Review”, May–June: 12–21.
- Kanwar, Vik. 2011. *Post-human Humanitarian Law: The Law of War in the Age of Robotic Weapons*, „Harvard National Security Journal”, 2: 616–628.
- Kaplan, Fred. 2013. *World as Free-Fire Zone*, „MIT Technology Review”, 116(4): 37–45.
- Keegan, John. 1993. *History of Warfare*, London: Vintage.
- Khun, Thomas Samuel. 1968. *Struktura rewolucji naukowych*, Warszawa: PWN.

- Kim, Alan, Brandon Wampler, James Goppert et al. 2015. *Cyber Attack Vulnerabilities Analysis for Unmanned Aerial Systems*, American Institute of Aeronautics and Astronautics, <https://engineering.purdue.edu/HSL/uploads/papers/cybersecurity/cyber-attack-lit-review.pdf> [15.02.2016].
- Kopeć, Rafał. 2014. *Rewolucja w sprawach wojskowych w kontekście zachodniego sposobu prowadzenia wojen*, „Kultura i Polityka”, 16: 64–80.
- Krushan, Armin. 2009. *Killer Robots. Legality and Ethicality of Autonomous Weapons*, Farnham: Ashgate Publishing.
- Kuptel, Artur i Andrew Williams (red.). 2014. *Multinational Capability Development Campaign (MCDC) 2013–2014. Role of Autonomous Systems in Gaining Operational Access*, Norfolk: Supreme Allied Commander Transformation HQ.
- Lin, Patrick, George Bekey i Keith Abney. 2008. *Autonomous Military Robotics: Risk, Ethics, and Design*, Washington: US Department of Navy, Office of Naval Research.
- Lin, Patrick, Keith Abney i George Bekey. 2013. *Ethics, War, and Robots*, w: Ronald L. Sandler (red.), *Ethics and Emerging Technologies*, New York: Palgrave MacMillan.
- Luttwak, Edward N. 1995. *Toward Post-Heroic Warfare: The Obsolescence of Total War*, „Foreign Affairs”, 74(3): 109–123.
- Magnuson, Stew. 2009. *Robo Ethics. Debate over rules, legality of robots on the battlefield lagging, expert say*, „National Defence”, 11: 28–29.
- Marra, William C. i Sonia K. McNeil. 2012. *Understanding „the Loop”: Regulating the Next Generation of War Machines*, „Harvard Journal of Law & Public Policy”, 36: 1139–1185.
- Meadows, Mark S. 2011. *We, Robot. Skywalker Hand, Blade Runners, Iron Man, Slutbots, and How Fiction Became Fact*, Guilford: Lyons Press.
- Morningstar, James K. 1997. *Technologies, Doctrine, and Organization for RMA*, „Joint Force Quarterly”, Spring: 37–43.
- Murphy, Richard i Afsheen J. Radsan. 2009. *Due Process and Targeted Killing of Terrorists*, „Cardozo Law Review”, 31(2): 405–450.
- Olson, Joshua i Muhammad Rashid. 2013. *Modern Drone Warfare: An Ethical Analysis*, American Society for Engineering Education Southwest Section Conference, <http://se.asee.org/proceedings/ASEE2013/Papers2013/157.PDF> [19.02.2016].
- Pandolfini, Bruce. 1997. *Kasparov and Deep Blue: The Historic Chess Match Between Man and Machine*, New York: Fireside.
- Parker, Geoffrey. 2008. *Historia sztuki wojennej. Od starożytności do czasów współczesnych*, Warszawa: Książka i Wiedza.
- Puckett, Christopher B. 2004. *In this era of „smart weapon”. Is a state under an international legal obligation to use precision-guided technology in armed conflict*, „Emory International Law Review”, 18(2): 645–723.
- Roberts, Michael. 1956. *The Military Revolution, 1560–1660: An Inaugural Lecture Delivered before the Queen’s University of Belfast*, Belfast: Boyd.
- Scharre, Paul. 2014. *Robotics on the Battlefield. Part I: Range, Persistence and Daring*, Washington: Center for New America Security.
- Schöndorf, Roy S. 2004. *Extra-State Armed Conflicts: Is There a Need for a New Legal Regime?*, „New York University Journal of International Law and Politics”, 37(1): 1–78.

- Searle, John R. 1980. *Minds, Brains, and Programs*, „The Behavioral and Brain Sciences” 1980, 3: 417–424.
- Sharkey, Noel E. 2010. *Saying ‘No!’ to Lethal Autonomous Targeting*, „Journal of Military Ethics”, 9(4): 369–383.
- Sharkey, Noel E. 2012. *The inevitability of autonomous robot warfare*, „International Review of the Red Cross”, 94(886): 787–799.
- Shaw, Martin. 2002. *Risk-transfer militarism, small massacres and the historic legitimacy of war*, „International Relations”, 16(3): 343–360.
- Simons, Geoff. 2011. *Drone Diplomacy*, „The Link”, 44(2): 3–13.
- Singer, P.W. 2010. *Wired for War. The Robotics Revolution and Conflict in the 21<sup>st</sup> Century*, New York: Penguin Books.
- Stout, Martha. 2005. *The Sociopath Next Door. The Ruthless Versus the Rest of Us*, New York: Broadway Books.
- Ślipko, Tadeusz. 2004. *Deontologia. Jej dzieje i miejsce w systemach etycznych*, „Studia Philosophiae Christianae”, 40(2): 83–97.
- Wallace, Lane. 2009. *Remote Control: Flying a Predator*, „Flying Magazine”, <http://www.flyingmag.com/pilot-reports/turboprops/remote-control-flying-predator> [20.03.2016].
- Windeck, Antoine. 2014. *Preface*, w: Ronan Doaré, Didier Danet, Jean-Paul Hanon i Gérard de Boisboissel (red.), *Robots on the Battlefield, Contemporary Issues and Implications for the Future*, Fort Leavenworth: Combat Studies Institute Press, s. V–VI.

#### THE ETHICS OF COMBAT ROBOTS

The application of autonomous combat systems is one of the most promising directions of military technology development, heralding a revolution in waging wars (the so-called “robolution”). Robots present a set of advantages over human soldiers. Interestingly, they are related not only to the sphere of operational efficiency, but also, potentially, to the sphere of ethics. However, a great deal of technological and ethical dilemmas should be resolved when using combat machines being robots *par excellence*: capable of operating autonomously (including making life and death decisions).

The article focuses on the challenges and possible solutions in building the artificial morality of autonomous combat systems. There are two development approaches – top-down and bottom-up – reflecting two ethical stances: rigorism and situationism.

Keywords: robolution, combat robot, autonomy roboethics, artificial morality