ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2017
Seria: ORGANIZACJA I ZARZĄDZANIE z. 110 Nr kol. 1985

Maciej ZAJĄC, Wojciech J. BOBER
University of Warsaw
e-mail: mzajac@uw.edu.pl, wjbober@uw.edu.pl

AUTONOMOUS MILITARY ROBOTS – CRITICAL TOPOGRAPHY OF POSSIBLE REACTIONS TO UPCOMING TECHNOLOGICAL REVOLUTION

Abstract. Recent technological advances make the prospect of partially or fully autonomous combat machines remaking the battlefields of the next decade very likely. The paper explores the space of possible reactions to such developments by the international community. We discuss three such families of approaches – an attempt at a comprehensive global ban; strict control regime – employment and strict regulation of Autonomous Military Robots by present-day global military powers coupled with vigorous non-proliferation efforts aimed at other countries and entities, akin to the management of nuclear weapons dynamic; and the laissez faire approach, i.e., exerting no effort to thwart the global proliferation of combat robots, whether military-grade or home-made, with all the consequences of such inaction. We conclude that while all these modes of reaction are burdened with very significant risks, vigorous action combining the elements of the global ban and control regime approaches is morally obligatory yet requires large conceptual effort to be undertaken on its behalf to be sufficiently effective.

Keywords: technical development, autonomous battle machines, Autonomous Military Robots

AUTONOMICZNE ROBOTY BOJOWE: MOŻLIWE REAKCJE NA NADCHODZĄCĄ REWOLUCJĘ TECHNICZNĄ

Streszczenie. Rozwój techniczny z ostatnich lat sprawia, że bardzo prawdopodobne stało się powstanie częściowo lub w pełni autonomicznych maszyn bojowych, które zmienią obraz pola bitwy w nadchodzącej dekadzie. W artykule zajęto się zakresem możliwych reakcji społeczności międzynarodowej na takie kierunki rozwoju. Omówiono trzy podejścia: próbę całkowitego zakazu, przyjęcie ścisłej kontroli (zastosowanie ścisłych regulacji w kwestii autonomicznych robotów bojowych przez obecne potęgi militarne wraz z przyjęciem polityki przeciwdziałania rozprzestrzenianiu, podobne do postępowania z bronią nuklearną) i podejście liberalistyczne, tj. niepodejmowa-

nie żadnych wysiłków w celu ograniczenia rozprzestrzeniania robotów bojowych, zarówno przez armie czy inne organizacje, ze wszystkimi konsekwencjami takiego zaniechania. Dochodzimy do wniosku, że jakkolwiek wszystkie te rodzaje reakcji obarczone są znacznym ryzykiem, moralnie pożądane jest podejście łączące elementy całkowitego zakazu ze ścisłą kontrolą, chociaż tutaj osiągnięcie sukcesu wymagałoby podjęcia zaawansowanych, skutecznych wysiłków badawczych.

Słowa kluczowe: rozwój techniczny, autonomiczne maszyny bojowe, Autonomiczne Roboty Wojskowe

1. Introduction

Autonomous Military Robots (AMRs) are, as we write these words in July 2017, a battlefield reality. While not yet fielded in large numbers in unstructured environments such as forests and cities, they have been placed in fortified border zones (Velez-Green 2015), flown off aircraft carriers (Holmes 2015) and employed on busy commercial waterways (Freedberg 2014)[1]. All that is already true for fully autonomous platforms; semi-autonomous systems and fire-and-forget munitions have been with around for decades. Still, the truly significant quantitative and qualitative leaps in autonomous weaponry are yet to be experienced. Before we analyze possible ways of shaping the incoming revolution in military technology, we need to demonstrate that mass introduction of AMRs would indeed constitute a revolution – and that, absent very significant international action, such revolution is inevitable.

For the purposes of this paper, we will define AMRs as all military systems capable of autonomously selecting and incapacitating human and/or material targets for at least an hour after being released by a human operator or by a vehicle operated, directly or remotely, by a human[2]. This definition separates AMRs from fire-and-forget munitions in a somewhat arbitrary way, but such a separation, although called for, is a matter of degree in self-sustainability and sophistication of autonomous targeting mechanisms. The difference between standard close-quarter air-to-air heat-seeking missile, targeting the closest source of heat for a period of seconds or minutes and a pilot-less, autonomous air superiority fighter targeting all Su-27s it detects within the duration of a three-hour patrol is profound, but nonetheless the exact threshold we set on the continuum between these two pieces of

[1] Although they have been armed and allowed to use their armament only in the first case.

[2] Target selection means that the machine attacks only a small region of the overall space within its striking range, unlike a mine, that indiscriminately brings destruction within the whole radius of its reach. Thus mines, mechanical traps, dams rigged with explosives, poisoned wells and other vehicles of indiscriminate, passive delayed killing do not fall under our definition.

hardware must be arbitrary. Our discussion will be aimed towards the higher end of that spectrum[3] – let the example of pilot-less interceptor be a reference point throughout.

For the introduction of so defined AMRs to cause a revolution in warfare – a revolution comparable to the appearance of firearms or nuclear weapons, and that is the scale of impact we predict – these platforms would need to outclass their near competitors on one of the three dimensions: casualty risk reduction, cost and performance. Given the trajectory of technological progress, AMRs should not only be compared against human soldiers, but also against a force consisting mostly of remotely-controlled, man-in-the-loop combat vehicles, or simply drones. Let us, therefore, evaluate three platforms – manned fighter featuring cutting edge sensors and software; same fighter remotely piloted by a human operator[4] from a control station in the rear-area bunker; and a fully autonomous fighter differing from the previous two only by the system commanding it being not a human brain but a specially designed piece of software implemented on computer embedded within its body.

In terms of casualty risk reduction, AMR fighter offers a clear and decisive advantage over manned counterpart – but so does the remotely-piloted craft. Both AMRs and drones completely protect their users from taking casualties and the political actors behind the war from having to explain the worth of the sacrifice. The difference between the two classes consists mainly in exposing the drone operators[5] to combat stress and consequent mental health issues. As tragic as they are, in current practice mental health injuries are not treated on par with physical injuries, death or capture by the enemy. Given that, AMR fighter would not be regarded as superior to the drone competitor in reducing casualty risk, and certainly not to a degree that would warrant completely phasing out the latter. Still, in the first of three categories, AMR at least holds its ground against the drone and outcompetes the manned alternative.

The difference between the price-tag of the three platforms would come mainly from the gap between the cost of training and compensation of the human pilots and that of developing an AI system capable of at least equaling their performance. As a fully functional combat fighter AMR has not yet been developed, an exact R&D figure cannot be provided. However, two different figures may give us a fairly clear picture: the entire development of X-47B fully autonomous platform cost, according to naval-technology.com[6], 813 million dollars – the sum

[3] Which does not mean discussing only high-end weaponry – our reasoning is equally valid for, and takes into account, a home-made mini-drone designed to slash at civilians in a street with a five-inch scalpel. For policy-relevant differentiation between high-end, low-end and home-made AMRs see (Zając 2017).

[4] Advances in technology offer the prospect of battlefield augmentation, such as performance enhancing drugs or exoskeletons, being developed to aid human warfighters. We do not discount the possibility that such augmented persons will possess greatly extended capabilities. But within the time framework of possible AMR fielding, we do not expect them to exceed the fundamental limitations of human body such as size, speed of reaction, need for food, water and sleep and vulnerability to contusions.

[5] For the sake of clarity we refer to humans in control of drone aircraft as 'drone operators', while aware that the skill sets required of such personnel are close to identical to that of manned platform pilots, and that historically most of the actual drone operators were combat proven fliers switching between the two roles.

[6] Available at http://www.naval-technology.com/projects/x-47b-unmanned-combat-air-system-carrier-ucas/

including creating and producing hardware of the platform, a lion's share of the overall program cost. The development of revolutionary software architecture for ALPHA, AI capable of reliably besting beating even the most experienced human pilots in simulated air combat, has been done on a 200-thousand-dollar grant in one year (Ernst et al. 2016, Reilly 2016). Once implemented, such software requires upgrades and new iterations, yet it never deteriorates in performance and can be used to fly essentially unlimited number of aircraft. Copied onto multiple airframes and secure, offline hard drives, it is also virtually invulnerable to corruption or destruction (though not to theft and replication).

In contrast, the cost of compensation throughout the career and retirement of a manned-fighter pilot – the lower-bound estimate being, for ease of calculation, one million dollars – is only a fraction of the sum required to develop and maintain her skills, a process that requires flying actual aircraft, generating fuel, maintenance and platform deterioration costs. According to data obtained in 2012 by Time Magazine[7], an hour of flight of a current high-end fighter craft costs from 40 to 70 thousand dollars, depending on the model; that means that the cost of even twenty hours of training exceeds one million dollars. Training an expert pilot takes several hundred hours, with the total cost reaching above ten million dollars. Still, even given the best retention policies, human pilots eventually retire and the whole process must begin anew. It is the case for both manned aircraft and drone pilots, although the latter may theoretically be able to retire at a later stage of their lives. Thus, fielding AMR fighters would not only pay for itself, with development costs being more than set off by savings on recruitment, training, compensation and retention, but would to a force with uniformly high skill level that is invulnerable to attrition or battlefield loses. Within the category of cost, employing AI pilots is, therefore, clearly much cheaper than employing humans, whether they operate the fighter directly or remotely. Note that this conclusion is ever more supported the less complicated the hardware used by human operators is; as the cost of a rifle is only a tiny fraction of the cost of training and employing the rifleman wielding it, the benefits of fielding a robot rifleman have the potential for much more spectacular cost reduction, at least when measured as proportion of the current costs.

The last parameter to be compared is performance. Autonomously taking off and landing on an aircraft carrier, X-47B has proven itself to have mastered a skill believed to be threshold of competency for human pilots. As far as software itself is concerned, combat fighter AI has already performed better than top human pilots. The ALPHA system, based on Genetic Fuzzy Tree methodology (Ernst et al. 2016), has, in the words of retired USAF colonel Gene Lee "seemed to be aware of my intentions and reacting instantly to my changes in flight and my missile deployment. It knew how to defeat the shot I was taking. It moved instantly between defensive and offensive actions as needed" (Reilly 2016). During repeated sessions in the realistic air combat simulator used for training fighter pilots, the veteran Lee got eventually

[7] Available at http://nation.time.com/2013/04/02/costly-flight-hours/, full data base content linked to in the body of the article.

shot down in every encounter, while never scoring a kill himself. “I go home washed out. I'm tired drained and mentally exhausted”, he told the reporter.

Implementing such software into an actual fighter plane is still far away. Yet the successes achieved by X-47B in terms of operational awareness and movement through space, and by the ALPHA team in terms of handling enormous quantities of data by breaking them down into sub-tasks placing only moderate amount of demand on the platform's computational power demonstrate that the most daunting challenges have already been surpassed. Having reached this level of maturity, fighter AI can mercilessly exploit all the advantages it holds over the human pilots of drones and especially manned aircraft.

To begin with, AMR aircraft contains no human body. It may maneuver at g forces that would kill a human, sustain hits that would incapacitate a human pilot, take shapes that would be impossible for manned aircraft (Committee on Autonomous Vehicles 2005, p. 137). It is never tired, angry or scared, and it does not have to prize its own survival at all, creating much greater flexibility for the human tactical commander and eliminating the need for extremely dangerous and risky combat search&rescue missions. It does not need to eat, drink and sleep.

All this, with the exception of insusceptibility to emotion, can be said of a drone operator, or at least about a team of them. Yet the edge held by an AMR is much greater than that. It is free not only of human bodily vulnerabilities, but also of the fundamental limitations on the speed of information processing, computational power, memory, and, as we already mentioned, ability to gain, retain and share skills and experience.

Transcendence of the limitations enumerated above translate into reflexes, survivability, decision competence and ease of acquiring it that could potentially get orders of magnitude better than those of the most talented and experienced human pilots. Humans had not evolved to be pilots or, generally, soldiers in a modern war. It is not at all surprising that machines and software meticulously designed with the single aim of warfighting excellence would surpass our bodies and minds, hastily and imperfectly adapted to such tasks. War is, indeed, an inhuman endeavor.

As we have demonstrated, AMRs equal drones and surpass human soldiers in reduction of casualty risk, are significantly cheaper than drone or human forces, and promise performance orders of magnitude better than the performance of top human warfighters or drone operators by being free of limitations hardwired into human biology. The cost reduction, or significant improvement in one specific area of performance – ability to break tight formations in case of shock cavalry, range/ease of use/psychological effect in case of firearms, lethality in case of nuclear weapons – are known to have not only triggered revolutions in military tactics and strategy, but also to lead to broad repercussions across all other area of human life. The AMRs offer extremely significant gains in ALL areas of performance, potentially confronting us with most profound change in the history of military technology. The rest of this paper will be

devoted to critical examination of the three possible strategies the international community may adopt towards the promises, challenges and dangers inherent in this revolution[8].

2. Laissez Faire Approach

The first approach to the incoming AMR revolution is the Laissez Faire Approach (LFA). Adoption of LFA would simply mean that individual states and their militaries, as well as the international community of nations would take no additional action to ensure that development and fielding of AMRs would not adversely affect the moral character of military operations, that is, their compliance with requirements of ad bellum and in bello justice. On one hand, LFA is a straw-man position – even those who do not believe AMRs to be morally problematic and welcome their development as potentially adding moral value (Anderson & Waxman 2013, Canning 2008, Kershnar 2013, Lucas 2013, Zając 2017) do not hold such an outcome to be automatic or easily achievable. All these authors point to specific challenges standing in the way of creating morally and legally permissible AMRs (challenges that have not been overcome yet), with George Lucas Jr. considering denial or ignorance of such challenges to constitute criminal negligence (Lucas 2013, pp. 227-228). On the other hand, LFA is a default scenario – the course events will take absent an intervention on some level. Therefore, serious discussion of LFA is due before we move to analyzing other approaches.

Let's first look at the Problem of Compliance with in bello justice, that is, making sure all the harm and destruction inflicted by the use of AMRs is proportional to the goals achieved through it and truly necessary for achieving such goals. It could be argued that the weapons' engineers, even if aiming only for ultimate battlefield effectiveness, would produce machines much more capable of compliance with in bello justice than human soldiers. After all, improving features such as precision or target detection and discrimination capability gives one a clear combat advantage. That is, however, a flawed assumption. The degree of precision or situational awareness necessary for destruction of the enemy force in usually lesser than that required for sparing civilians caught in the midst of the battle or enemy personnel willing to surrender. Moreover, equaling human potential cannot be the end of the effort towards realizing the true potential of AMRs regarding humane conduct of war. However, industry executives and procurement offices, especially if an arms race dynamic sets in, are going to put pressure on quick and inexpensive development with only minimal, not optimal, compliance standards met. Adding the fact that there may exist emergent problems with compliance that we cannot predict now, its seems clear that absence of additional pressure on

[8] We will limit our analysis to tangible harms and benefits ARM introduction may bring. More esoteric concerns, such as those about warfare being made less chivalrous, combatant dignity being detracted from or arms being inherently wrong independently of their ability to fulfill the requirements of the Laws of War will not be discussed (for an argument stating that such concerns are entirely groundless Kershnar [2013]).

robust and much more restrictive compliance standards will result in a missed opportunity for making warfare orders of magnitude more humane than it is today. To illustrate – one frequently floated proposal for the rules of entanglement placed on combat robots is having them target weapon systems rather than their human users, and, provided that cannot accomplish the mission, engaging personnel in less lethal way, for example by shooting at their extremities rather than effecting much more traumatic torso or head wounds. Given the advantage AMRs would have over any human force, and the precision with which they would be able to target human opponents, such a scenario seems achievable. Still, it is difficult to imagine such instructions programmed into these robots for narrowly defined military purposes. Extra considerations – political and moral ones – would need to be brought to bear, and it is exactly such a debate about new standards for in bello justice that would be absent if LFA was adopted. Michael Walzer's principle of double intention requires additional capacity to be translated into additional benefits for civilians, and so would be violated by LFA (Walzer 1977, pp. 155-159).

Ius in bello compliance problems are, comparatively, the least significant ones. The Arms Race Problem is cause for much more concern. As with any truly transformational military technology, even a mere prospect of it being acquired by one power triggers a rush to develop it by its rivals. Temporary dominance gained by the winner of the race does not usually lead to conflict, as it may be offset by competitors in some other domain. For example, the threat posed to China by unprecedented American dominance on high seas is compensated for by Chinese advantage in littoral warfare. However, AMR technology offers the states that effectively pursue it a prospect of dominance across the entire conventional spectrum (and in the long term, of game-changing advance in anti-nuclear capability)[9]. Such a rapid shift in the balance of power may lead to a preventive strike by an adversary who presumes his position to be greatly weakened. Given that the most likely beneficiaries and losers in such breakthroughs are nuclear powers, the stakes humanity has in checking the unfolding of such conflicts are extremely high[10]. To propose that there should not be a global treaty regime regulating such progress, nor bilateral treaties between individual powers, nor even talks, consultations and assurances between them is to advocate for multiple uncontrolled upsets to the global balance of power at roughly the same time. No moral, political or strategic thinker would do so, and so LFA seems clearly bankrupt as a way of dealing with AMR revolution.

This holds even before we consider the effect adopting LFA would have on general compliance with ius ad bellum – justice regarding starting and continuing wars. We do not believe that AMR revolution would necessarily lead to greatly increased incidence of unjust conflicts, or that the harm generated by such conflicts would outweigh the harm avoided by

[9] Think of what placing thousands of ARM fighters in the sky at all times could do for ballistic missile interception.

[10] Similar conditions seem to hold with regard to the research and development of strong Artificial Intelligence, and studying either problem seems to be informative in connection to the other – see Bostrom (2014), pp. 84-85.

increased ability of some agents to engage in just wars[11]. Still, we admit that most regimes existing today are not capable of conducting just war, with many of them failing to meet Brian Orend's criteria for a minimally just state (Orend 2006), or John Rawls's criteria for a well-ordered state (Rawls 1999, pp. 83-84). Given this reality, unchecked proliferation of extremely effective weaponry is by no means an acceptable outcome. This is compounded by the fact that universal proliferation of AMRs would make them accessible not only to states or already non-state armed groups, but also to a broad range of actors currently incapable of effecting military-grade violence. Possession of skilled manpower no longer being necessary for operating a military force, corporations, obscure groups or even lone-wolf fanatics would be able to wage mass violence without directly endangering themselves in any way. The prospect of petty dictators fortifying their power with unconditionally loyal robot goons, or a single radicalized explosive specialist turning swarms of delivery drones into flying bombs appeals to no one. LFA is distinctively unattractive from each and any perspective, and it is probable that it is LFA the most fierce opponents of AMR technology have in mind when calling for a global ban on such weapons. Let us in turn take a look at this approach, seemingly simple and attractive, yet ultimately also burdened with unacceptable flaws.

3. Global Ban Approach

In its most radical form, Global Ban Approach (GBA) on automated weapons means simply that all the R&D activities aiming at their construction are stopped worldwide and no such armaments are introduced by the military. At the first glance, it may seem that the GBA on AMRs' use is not very probable, yet we must note current attempts at introducing it. In December 2016, The Fifth Review Conference of the Convention on Conventional Weapons at the United Nations in Geneva set the course toward a ban on "killer robots" (as autonomous weapons are often called) with 19 countries calling for a complete ban (Algeria, Argentina, Bolivia, Chile, Costa Rica, Cuba, Ecuador, Egypt, Ghana, Guatemala, the Holy See, Mexico, Nicaragua, Pakistan, Panama, Peru, the State of Palestine, Venezuela and Zimbabwe[12]) and further, including China, seeing the need for international instruments to control autonomous weapons.[13] However, the future of this initiative is not clear as plans scheduled for 2017, including important meeting in August, were suspended due to financial problems. Besides, it should be noted that the countries calling for GBA include no major military power and only few actors involved in current conflicts or placed near the territories where armed drones or other similar weapons have already been used, as the Near East or Northern Africa (Algeria,

[11] Occurrence of just wars being, *ex definitione*, morally preferable to their non-occurrence.
[12] Campaign to Stop Killer Robots (2016)
[13] Human Rights Watch (2016) https://www.hrw.org/news/2016/12/16/un-key-action-killer-robots

Egypt, Pakistan and the State of Palestine); majority of the countries calling for the global ban is placed in South America, far from major conflicts.

GBA is also promoted by organizations such as Campaign to Stop Killer Robots.[14] If enforced, it would prevent the development, production, and use of weapons that do not exist yet. It would go in line with considerations of such philosophers as Hans Jonas or, more recently, as Michael Walzer. Jonas developed a kind of "precautionary principle" already in 1970s; according to him, we should give the bad prognosis precedence over the good one with caution being placed at the core of moral action.[15] On the other hand, precautionary principle, much discussed in recent times and included into international law, excludes following uncertain or dangerous paths of conduct, especially those impacting the environment.[16] If applied to the problem of AMRs, it would favor suspending the development of such weapons given that consequences of such development would be either hard to predict or dangerous. Walzer, whose principle of double intention has already been mentioned, in his commentary on Kosovo intervention, claims boldly that "You can't kill unless you are prepared to die."[17] Adoption of such a principle would certainly delegitimize all uses of AMRs that could lead to human death[18]. It would go in line with claim made by Joseph Weizenbaum already in 1960s that we should not leave certain matters that involve compassion and judgement to robots or AI. Killing people certainly would belong to that group. Other prominent and visible proponents of GBA on AMRs include Sharkey (2010) and Sparrow (2007), as well as Strawser (2010), who, interestingly, simultaneously believes the use of properly piloted remotely controlled platforms to be morally mandatory.

Attempts at excluding certain kinds of weaponry are far from being a new invention. But it should be noted that historical bans, as those imposed on crossbow in the Middle Ages in Europe or on firearms in Japan, did not prove to be a great success, at least in a longer perspective. At present, international conventions ban generally two kinds of weapons of mass destruction (biological and chemical) as well as use of several conventional weapons (including, among others, anti-personnel mines or blinding laser weapons). But no kind of conventional weapons is banned completely. Besides, all the questioned armaments were constructed and used (with perhaps the exception of blinding lasers) on the battlefield before respective bans have been passed. Banning AMRs both as a whole class of weaponry and

[14] See: https://www.stopkillerrobots.org/
[15] Jonas 1984, 37-38.
[16] See e.g. Luján & Todt 2012.
[17] McMahan 2013, p. xi.
[18] Such an interpretation of the Principle of Double Intention may be rendered invalid by the very technology in question. Walzer has wrote the quoted passage in the context of the bombing of Serbia, in which endangering the combatants via lowering of flight ceilings would have offered better protection to civilians. However, in the near future, this need not be the case. It may well be that using AMR instead of humans will offer the civilians much better protection – the goal of Walzer's prescription. If there was a less pragmatic component to Walzer's views on the matter – and there probably is, given the general sentiment prevailing in chapters 2 and 3 of Walzer (1977) – in the case of AMRs the pragmatic and deontic components of Walzer's justification for his views may be at odds, rather than reinforcing each other.

before their construction would be, therefore, certainly an exceptional case in human history and unlike any other existing convention.

More decisive argument against GBA consists in the fact that enforcing the end of progress in the field of autonomous robots seems fully improbable. You need not believe in technological determinism to predict that such robots may prove to be useful in many areas far from combat, and, therefore, the development of respective AI and hardware seems both inevitable and beneficial. If once created, such technology may be easily transferred to the military given the fact that many remotely controlled robots are already in service (not to mention that the military is much involved in such developments) or could be adopted by, e.g., terrorists with unpredicted consequences. As P. W. Singer, a writer much skeptical about many aspects of the AMR revolution, has put it: "It is hard to imagine a future with any outright ban of autonomous technology, even in war. To do so is to imagine a world in which a military pilot is driven to his base by his robotic car, and then fights a battle in which all sides have agreed to use only older technologies" (Singer & Cole 2016).

It seems that at least development of "killer robots" programmed for "killing" other "killer robots" is not only unobjectionable from both moral and pragmatic point of view but, perhaps, also advisable. Moreover, military equipment often can serve purposes far from combat, e.g. in rescue missions, quite common not only at the battlefield but also in case of natural disasters. Even if there are inherently bad kinds of weaponry (as it is often claimed on nuclear weapons), certainly not all autonomous robots must fall into that category. Previously mentioned autonomous guns implemented in border zone between two Koreas, where no civilians are expected to appear, pose no major moral and legal objections (of course, if situation would change in such a way that civilians fleeing from the North could appear, the evaluation would change as well). Machines outperforming humans in certain aspects may cause less casualties, as we have already stated – a reason that always should be taken into consideration.

4. Enforced Restrictions Approach

GBA is not the only alternative to LFA. In the middle ground lies an option of enforcing certain restrictions on AMRs **without making them completely illegal**. Let us call this possibility enforced restrictions approach (ERA). AMRs' development and use is to be placed under control of internationally deliberated rules, similarly to other conventions governing different kinds of armaments and actions during the war.

Till this point, it seems that it is the most reasonable solution, given the inherent and irresolvable difficulties of both LFA and GBA. The problem is, however, the shape such conventions or restrictions should take. As we are just at the beginning of the revolution, it

may prove to lead us into many unpredicted – and, perhaps, unpredictable – outcomes. What we certainly can do is to learn the lesson drawn from the introduction of drones, which should be compared with autonomous weapons not only on economic grounds. It is not the case that we should model the conventions dealing with AMRs solely after moral problems that already appeared. With time, these problems may be aggravated or genuinely new problems may emerge, including those involving agency and moral status of AI[19]; rather we should not follow the laissez faire approach that accompanied introduction and early use of drones, with little attention on the part of human rights and humanitarian organizations (Alston 2012, p. 36). Already mentioned similarity of drones to at least part of planned AMRs provides evidence for moral and legal problems connected to drones reappearing in the case of autonomous weapons as well.

The most important moral and legal problems with drones include, among others, challenges to understanding the size and extension of a combat zone, of who counts as combatant or what constitutes the conditions of surrender; involvement of this new technology simply challenged our traditional understanding of such phenomena, calling for both legal and ethical analyses. In recent years, much focus was placed on the problem of the so-called "targeted killing" performed with use of drones[20]. Such operations took place in areas distant from regular combat, were at times directed towards persons whose actual engagement in conflict could be questionable and in a way that might be regarded as extrajudicial execution rather than regular combat. Problems of surrendering combatants and attacks on those helping already injured persons have been reported as well. Certainly, the world evolves and our understanding of factual and moral matters evolves as well. We do not postulate that if a kind of armament challenges our understanding of what is permissible in war then it should be banned; but each occurrence of such a challenge should be investigated so that we could arrive at consistent set of rules, preserving us from sinking into the state of barbarism (Walzer 1977, p. 44-46). Such as set of rules would need subsequent implementation, and engineering task to be carefully performed on an industry scale, requiring an effort to raise awareness and upgrade the professional ethics of engineering and management (Lucas 2013). While some conceptual attempts at this seem to be promising (Arkin, Ulam & Duncan 2009), ERA would require significant modifications of R&D, testing, fielding, monitoring and war crime prosecution even at the level of a single country, military branch or armaments company (Lin, Bekey & Abney 2008, p. 73-86).

The problems to be encountered at national and sub-national level are merely one class of problems connected to ERA. The other is that of international arrangements to be put in place – arrangements hardly restricted to corrections or expansions of existing legal provisions

[19] These matters have been widely discussed during last decades, both from moral and legal point of view. We provide no references, as the problem is marginal to present considerations, but proponents of ascribing moral status to robots are many.
[20] See e.g. Whetham 2013.

regarding *ad bellum* or *in bello* justice. Stated simply, the advantage provided by employment of AMRs in combat as well as in massacre is such that states could and will react drastically to a real or perceived "AMR gap". Bi- or multilateral treaty frameworks may need to be established to check the incoming arms race – a feat that may prove very difficult given the existence within present day international environment of many other toxic dynamics. On top of that, the continued existence of both rogue states and failed states presents the international community with the dilemma of trying to block the proliferation of AMRs into the territories over which it does not exercise control through soft power, and where hard power or threats of its use have a spotty record at best. The situation is further complicated by the fact that many of said areas are considered client states or buffer zones by some of the powers key to any effective and lasting non-proliferation effort.

The nuclear non-proliferation regime seems to suggest a framework – chief global powers with access to the most advanced weaponry: 1) keeping each other in check, ready to answer any substantial move to break the balance of technological prowess with a similarly effective response, 2) providing military support and guarantees to countries under their "umbrella" in amount sufficing to offset these countries non-pursuit of their own AMRs[21] and 3) jointly intervening, or at least allowing other powers' intervention, in such rogue or failed states that fail to become a part of the "umbrella" arrangement with any AMR power and do try to develop or proliferate AMRs, or host groups performing such actions.

Even the establishment of such a framework – a difficult, long and probably violent process that it would be – would not stop the threat posed by civilian robots, such as self-driving cars or delivery drones, being weaponized and used with deadly affect by individuals having loose or none ties to states or large non-state groups. The potential of a single individual to inflict mass casualties while exposing himself to no direct harm is another so far unchecked downside of the robotic revolution, and given how politically destabilizing terrorism may be at home and abroad – think only of the aftermath of 9/11attacks – damage inflicted by such actors may be far greater than the harm to the victims who have suffered directly.

5. Conclusion

The already ongoing change in military technology, connected with development and deployment of Autonomous Military Robots, fully deserves to be called revolutionary. So will be its impact. In the first part of this paper, we have explained that this revolution is a consequence of the well-established trajectory of modern technology and certain fundamental

[21] Or perhaps by providing them with a limited number of AMRs in exchange for non-development of technological and industrial capacity for domestic production and proliferation.

facts about the limitations of human body and cognition. We have critically discussed three broad approaches that could be taken towards this wave of change by the international community. The Laissez Faire Approach, that of non-intervention and permission for matters to take their course unobstructed, we have deemed deeply unsatisfactory and in some of its forms and aspects tantamount to criminal negligence. Given the AMR technology's tremendous potential for causing both good and evil, inaction would be as staggeringly immoral as it would be imprudent and unstatesmanlike. A second approach, the one of Global Ban on AMRs, is in our opinion completely implausible. It would require countries and their militaries to forgo very substantial advantages and ignore opportunities to use the AMR technology for improved compliance with the Laws of War and the spirit of Just War Theory, while simultaneously enforcing an absurd and growing disparity between the quality of devices used for civilian and military purposes. We fear that well-intentioned attempts to implement the Global Ban, while doomed to fail, have a potential for causing very significant harm on their own.

The approach we advocate for is that of Enforced Restrictions – strict and global arms control regime ensuring each and every AMR platform's compliance with the Laws of War and restricting their proliferation as much as possible, akin to the universally accepted nuclear weapons policy. In that way the harm of these terribly efficient weapons being used by rogue states or tribal warlords could be avoided. Enforced Restrictions Approach is, as we admit, fraught with dangers, and even if executed well – a challenge in itself – does not resolve all moral and policy issues arising from the automation of military forces. Still, it is the only approach that offers any chance of avoiding the horrors of unconstrained "killer robots" finding their way under the command of ill-intentioned handlers, while harnessing some of the potential of the new technology for alleviating the harms of war. The success of this approach, we stress, is not certain, and may even be unlikely – yet it is our best shot, and we are morally bound to take it. Ethicists, technologists, scientists and policymakers alike should work together to devise and implement the best possible version of such approach. Given the scarcity of actual, fleshed-out solutions at this point in time, the need for further research is self-evident.

Bibliography

1. Alston (2012) – Philip Alston, "Lethal Robotic Technologies: The Implications for Human Rights and International Humanitarian Law," Journal of Law, Information and Science, 21(2), pp. 35-60.
2. Anderson & Waxman (2013) – Kenneth Anderson & Matthew C. Waxman, "Law and Ethics for Autonomous Weapon Systems; Why a Ban Won't Work and How the Laws of

War Can", Stanford University, The Hoover Institution Jean Perkins Task Force on National Security and Law Essay Series, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2250126 [accessed: 25.04.2017]

3. Arkin, Ulam & Duncan (2009) – Ronald C. Arkin, Patrick Ulam & Brittany Duncan, "An Ethical Governor for Constraining Lethal Action in an Autonomous System", Georgia Institute of technology online publication, 2009, available at http://www.cc.gatech.edu/ai/robot-lab/online-publications/GIT-GVU-09-02.pdf [accessed: 25.04.2017].
4. Beauchamp & Savulescu (2013) – Zack Beauchamp & Julian Savulescu, "Robot Guardians; Teleoperated Combat Vehicles in Humanitarian Military intervention", Killing by Remote Control. The Ethics of an Unmanned Military, (eds.) B. Strawser & J. McMahan, Oxford, Oxford University Press, 106-125.
5. Bostrom (2014) – Nick Bostrom, "Superintelligence – Paths, Dangers, Strategies", Oxford, Oxford University Press 2014.
6. Campaign to Stop Killer Robots (2016) – "Country Views on Killer Robots," available at http://www.stopkillerrobots.org/wp-content/uploads/2013/03/KRC_CountryViews_13Dec2016.pdf [accessed: 1.07.2017]
7. Canning, (2008) – John Canning, "Weaponized Unmanned Systems: A Transformational Warfighting Opportunity, Government Roles in Making It Happen", Proceedings of Engineering the Total Ship (ETS), September 23-25. 2008, Falls Church, VA, available at http://www.sevenhorizons.org/docs/CanningWeaponizedunmannedsystems.pdf [accessed: 25.04.2017].
8. Committee on Autonomous Vehicles (2005) – Committee on Autonomous Vehicles in Support of Naval Operations National Research Council, "Autonomous Vehicles in Support of Naval Operations", Washington, DC: The National Academies Press, 2005
9. Cummings (2006) – Mary L. Cummings, "Automation and Accountability in Decision Support System Design", Journal of Technology Studies, Volume XXXII, nr1, winter 2006, pp. 23-31.
10. Ernst et al. (2016) – Ernest N, Carroll D, Schumacher C, Clark M, Cohen K, et al., "Genetic Fuzzy Based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions", March 22nd 2016, Journal of Defense Management 6:144.
11. Freedberg (2014) – Sidney J. Freeberg Jr., "Naval Drones 'Swarm', But Who Pulls the Trigger?", October 5th, 2014, available at http://breakingdefense.com/2014/10/who-pulls-trigger-for-new-navy-drone-swarm-boats/.
12. Freedberg (2016) – Sidney J. Freeberg Jr., "Marines Seek to Outnumber Enemies with Robots", October 25th, 2016, available at http://breakingdefense.com/2016/10/marines-seek-to-outnumber-enemies-with-robots/

13. Holmes (2015) - "The Mighty X-47B – Is it Really Time for Retirement?", National Interest May 6, available at http://nationalinterest.org/feature/the-mighty-x-47b-it-really-time-retirement-12818 [accessed: 25.04.2017].
14. Human Rights Watch (2016) – "UN: Key Action on 'Killer Robots': International Move Toward Possible Ban", December 16, 2016, available at: https://www.hrw.org/news/2016/12/16/un-key-action-killer-robots [accessed: 1.07.2017]
15. Jonas (1984) – Hans Jonas, The Imperative of Responsibility. In Search of an Ethics for the Technological Age, Chicago, Chicago University Press 1984.
16. Kershnar (2013) – Kershnar, S., 2013, "Autonomous Weapons Pose No Moral Problem", Killing by Remote Control. The Ethics of an Unmanned Military, (eds.) B. Strawser & J. McMahan, Oxford, Oxford University Press, 229-245.
17. Lin, Bekey & Abney (2008) – Patrick Lin, Paul Bekey & Keith Abney, "Autonomous Military Robotics: Risk, Ethics and Design", California Polytechnic State Univ San Luis Obispo, December 20th 2008, available at http://ethics.calpoly.edu/onr_report.pdf
18. Lucas (2013) – George Lucas Jr., "Engineering, Ethics and Industry: The Moral Challenges of Lethal Autonomy", Killing by Remote Control. The Ethics of an Unmanned Military, (eds.) B. Strawser & J. McMahan, Oxford, Oxford University Press, 211-228.
19. Luján & Todt (2012) – José Luis Luján and Oliver Todt, "Precaution: A taxonomy", Social Studies of Science, Vol. 42, No. 1 (February 2012), pp. 143-157.
20. McMahan (2013) – Jeff McMahan, "Foreword", Killing by Remote Control. The Ethics of an Unmanned Military, (eds.) B. Strawser & J. McMahan, Oxford, Oxford University Press, ix – xv.
21. Walsh, Tegmark et al. [2015] – Multiple Signatories, "Autonomous Weapons: An Open Letter From AI & Robotics Researchers", https://futureoflife.org/open-letter-autonomous-weapons/ [accessed: 25.04.2017].
22. Orend [2006] - B. Orend, „The Morality of War", Broadview Press, Peterborough, Ont. 2006.
23. Rawls [1999] – J. Rawls, "The Law of Peoples", Harvard University Press, Cambridge, Mass. 1999, second edition.
24. Reilly (2016) – M. B. Reilly, "Beyond Video Games: New Artificial Intelligence Beats Tactical Experts in Combat Simulation", University of Cincinnati Magazine, June 27th 2016, available at http://magazine.uc.edu/editors_picks/recent_features/alpha.html
25. Sharkey [2010] – Noel Sharkey, „Saying 'No!' to Lethal Autonomous Targeting", Journal of Military Ethics, 9, 4. (2010).
26. Singer & Cole [2016] – Peter W. Singer & August Cole, "Humans Can't Escape Killer Robots, but Humans Can Be Held Accountable for Them", Vice News, April 15th 2016, available at https://news.vice.com/article/killer-robots-autonomous-weapons-systems-and-accountability [accesssed: 11.07.2017].

27. Sparrow [2007] – Robert Sparrow, “Killer Robots”, Journal of Applied Philosophy 24, 1., (2007).
28. Strawser [2010] - Bradley Strawser, “Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles”, Journal of Military Ethics, 9 (4):342-368, 2010.
29. United States Department of Defense, 2012 – United States Department of Defense “Directive Number 3000.09; Subject: Autonomy in Weapon Systems”, available at http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf [accessed: 25.04.2017].
30. Velez-Green (2015) – Alexander Velez-Green, “The South Korean Sentry – A 'Killer Robot' To Prevent War”, Lawfare, available at https://www.lawfareblog.com/foreign-policy-essay-south-korean-sentry%E2%80%94-killer-robot-prevent-war [accessed: 25.04.2017].
31. Warrick (2017) – J. Warrick “Use of weaponized drones by ISIS spurs terrorism fears”, available at https://www.washingtonpost.com/world/national-security/use-of-weaponized-drones-by-isis-spurs-terrorism-fears/2017/02/21/9d83d51e-f382-11e6-8d72-263470bf0401_story.html?utm_term=.f8e1c6508bee [accessed: 25.04.2017].
32. Walzer (1977) – Michael Walzer, “Just and Unjust Wars”, Basic Books 1992, Second Edition.
33. Whetham (2013) - David Whetham, “Drones and Targeted Killing: Angels or Assassins?”, Killing by Remote Control. The Ethics of an Unmanned Military, (eds.) B. Strawser & J. McMahan, Oxford, Oxford University Press, 69-83.
34. Zając (2017) – Maciek Zając, “Infeasibility of the 'Global Ban' Approach to Autonomous Robots”, forthcoming in Proceedings of the 'Violence and Society' International Conference, Warsaw 2017.