Review Article

AI, Robotics and Ethics: A Dangerous Slippery-Slope of Unmanned Autonomous Weapons Use in Modern Warfare

Predrag T Tošić[1,2]*

[1]*Affiliate/Adjunct Faculty, Department of Mathematics & Statistics, Washington State University, Pullman, Washington, USA*

[2]*Independent AI and Data Science consultant and author*

*Corresponding author: Predrag T Tošić, Affiliate/Adjunct Faculty, Department of Mathematics & Statistics, Washington State University, Pullman, Washington, USA.

Abstract

Rapid advancements in AI and Robotics (including various types of autonomous unmanned vehicles) are changing almost every aspect of the world we live in. While holding a great promise to make this world a better place insofar as applications of advanced AI in health care, education, manufacturing, agriculture and other domains, AI-based weapons and their use in modern warfare, in current and future armed conflicts, can potentially be extremely devastating. A prominent group of AI researchers, technologists and other scientists and scholars penned an Open Letter in 2014-2015 about these great dangers to the human kind, in which they called for a broad international ban on AI-based weapons. We review the key elements and growing relevance of this Open Letter and especially the call for a ban, in the context of the ongoing war in Ukraine in which the world is witnessing a heavy use of unmanned drones and other autonomous, AI-enabled technologies by both sides in that large-scale armed conflict with high casualties among both the two militaries and the civilian population.

One-Sentence Summary: This short article outlines the growing ethical challenges and concerns stemming from "automation and robotization of modern warfare", and in particular the dangers from ever-growing use of AI and Robotics, such as aerial and ground unmanned vehicles and drones being heavily used in recent and ongoing conflicts in the former Soviet geopolitical sphere.

Keywords: Artificial Intelligence; Robotics; Unmanned vehicles; Drones; Ethics; Modern warfare

Abbreviations: AI: Artificial Intelligence

Introduction

It has been over 8 years now (as of the early 2023) since a large groups of technologists, scientists, AI experts, industry leaders from Google, Apple, Microsoft and other tech companies, university professors and others supported and signed an *open letter* voicing their concerns about the use of AI in contemporary and future warfare; and arguing in favor of a world-wide, United Nations supported and mandated, ban on such AI-based weapons. Among the original group of about 1,000 signatories were Elon Musk (of TESLA and more recently Twitter fame), Steve Wozniak (a Silicon Valley veteran and co-founder of Apple with Steve Jobs), and the late famous physicist Stephen Hawking [1]. The still-open list of signatures and the text of the original letter are maintained at website [2]. (This author, while not being a prominent individual like many of the names on that list, has recently decided to himself sign the call for a global ban on AI-based weapons, as the world is witnessing the ongoing carnage in Ukraine, including heavy use of drones and other AI/robotics/autonomous unmanned vehicle technologies and weaponry). The list of supporters of the Open

Letter and the ban proposed in it is currently (as of early January 2023) more than 34,000 names long, and keeps growing. What is the crux of the issues raised in that *open letter*, what is the broader relevance of those concerns and, very importantly, what are the odds that the call for an international ban on "AI weapons" would be heeded by the world community and the governments and militaries around the globe?

The letter states: "AI technology has reached a point where the deployment of [autonomous weapons] is – practically if not legally – feasible within years, not decades, and the stakes are high: <u>autonomous weapons have been described as the third revolution in warfare, after gunpowder and nuclear arms</u>" [emphasis added]. Per Guardian's article [1], the authors of the letter "argue that AI can be used to make the battlefield a safer place for military personnel, but that offensive weapons that operate on their own would lower the threshold of going to battle and result in <u>greater [overall] loss of human life</u>" (emphasis added). Recent armed conflicts, such as (relatively modest in scale) the conflict between Armenia and Azerbaijan over the region of Nagorno-Karabakh, and the ongoing, much larger in scale war in Ukraine involving the Russian Federation's military and Russian-backed rebels in parts of eastern Ukraine on one side and the Ukrainian Armed Forces (UAF) and some paramilitaries aligned with UAF on the other, have both involved a considerable use of unmanned drones and are the "real world examples" justifying the concerns raised in the Open Letter found at [2]. It is practically certain that these conflicts and the use of autonomous weapons and drones in them, especially with the large-scale, high-tech war in Ukraine with high casualties on both sides (including those inflicted by drones and other "AI-based" weapons), will be seen as "we told you so" by the historians of AI and warfare alike in the years and decades to come. Indeed, we can already argue that Musk, Hawking, Wozniak et al., with their letter originally penned in 2014 – 2015, have already been proven prophetic.

Timely Concerns on AI And Robotics in Modern Warfare

"The endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. The key question for humanity today is whether to start a global AI arms race or to prevent it from starting," the authors of the Open Letter argue [1]. Reference to Kalashnikovs warns that not banning AI weapons and robotized war-waging will result in the (likely, fairly near) future acceptance of such AI weapons as something casual and to be expected, ultimately leading to "the new normal", and even banality, of large-scale destruction and loss of life these weapons already are, and, should they be allowed to continue being developed and then deployed in future armed conflicts, will become even more capable of causing in the future. However, there is, at least in my reading of the Open Letter, another, more subtle, if only implied, reference; namely, possible parallels between an AI armed race (unless it is prevented early on, that is, now!) and how the nuclear armed race ensued at the end and in the aftermath of WW2, and subsequently throughout the Cold War. For example, per Guardian's article [1], Toby Walsh, professor of AI at the University

of New South Wales said: "We need to make a decision today that will shape our future and determine whether we follow a path of good. We support the call by a number of different humanitarian organisations for a UN ban on offensive autonomous weapons, similar to the recent ban on blinding lasers." Significant as it is, banning of blinding lasers dwarfs in comparison to the war-making potential of AI and Robotics broadly defined, and likewise how far-reaching would a broad ban on robotized warfare really be. Of course, as with a ban of any type of weapons or behavior in warfare, there is always the issue of how to enforce actual compliance with whatever international restrictions or bans on AI and Robotic weapons and war-waging might be imposed by the international organizations such as the United Nations.

With regards to the use of atomic/nuclear weapons, the only actual use in a war was the atomic bombs being dropped by the Allies (the United States) on two Japanese cities, ostensibly to bring about Japanese capitulation sooner rather than later; and thus saving human lives overall – or at least this has been the argument justifying the use of atomic bombs; and contrasting the loss of human life in that scenario as opposed to a "traditional" invasion of Japan's mainland by the US and western allies in order to bring Japan's capitulation, and the estimated casualty tolls on both the Americans/Allies and the Japanese in that scenario. Thankfully, the world has not seen the use of atomic and nuclear weapons since 1945 (other than in the context of testing such weapons). Authors and supporters of the letter on the use of AI in warfare are well aware of the history of the nuclear arms race; they indeed want to prevent such a race from starting in the context of AI weapons. However, it is highly questionable, noble as the intent of their Open Letter is (and, as not only a modest AI researcher but first and foremost as a human being and a Christian, I surely share those concerns as well as support the proposed ban!), how successful the calls for the ban will be today or in the foreseeable future, given the world of *realpolitik*. Some of the world's leading military powers, as well as leading nations in AI research (including democracies!), have already opposed an international ban on R&D on "killer robots" (and by extension, on other autonomous unmanned vehicles used in warfare); see, for example, [3]. So, what can be done to prevent the doomsday scenarios outlined in the Open Letter, and is there a political will and a realistic chance of accomplishing global international agreements that would reduce or ideally eliminate those doomsday scenarios related to AI-based, robotized warfare potentially leading to unprecedented large-scale destruction and loss of human life?

AI, Autonomous Weapons, Drones and The Moral Agency

This author has been interested in the concept of (autonomous) "agency" since being a graduate student in the early 2000's; as a part of a broader vision and related academic research focusing on AI-based systems that would not only assist humans in making decisions or executing various tasks, but would one day actually act fully autonomously (from human operators), including autonomous decision making and acting in various types of complex, partially observable and dynamic environments [4].

Page 2 of 4

While AI has tremendously advanced in the meantime, various types of autonomous unmanned vehicles already existed and were used in armed conflicts, as well as researched extensively at many government, private sector and academic labs, including by my PhD advisor's research group at the University of Illinois, at that time (around year 2000). In particular, autonomous *Unmanned Aerial Vehicles* (UAVs), commonly referred to as 'drones', as well as various types of ground unmanned vehicles have been used by several nations' armies/militaries, including an extensive use by the US military, in various armed conflicts in the 1990s and into the early years of the 21st century; with heavy investment into related R&D by the US Department of Defense (DoD) and similar agencies at other leading technology and military powers including the UK, Russia and China. Already by the year 2000, the United States government (and very likely several other governments, those with, if not quite comparable to the US, then certainly advanced with respect to the rest of the world technological, research and military capabilities) have been envisioning the use of, and heavily investing into related research and technology development related to, the next-generation of fully or nearly fully autonomous aerial, ground and underwater vehicles. The early adoption as well as envisioned future uses of such autonomous unmanned vehicles comprised both military and non-military applications, including for various types of exploration and surveillance in scenarios where sending human-operated vehicles would be potentially too dangerous or infeasible [5].

Admittedly, my early interests and research in 'agent autonomy' and autonomous action/decision-making of AI-based systems was not primarily inspired by ethical concerns, but rather by the engineering design as well as amenability to formal mathematical analysis of behavior of such systems. In that context, in work mentored and guided by my advisor prof. Gul Agha at the Computer Science Department at the University of Illinois, I discussed the dangers of "excessive antropomorphization" of AI-systems, insofar as attributing human-like qualities to a piece of software, an autonomous drone or a robot – properties such as beliefs, desires, intentions and so forth [6]. Such anthropomorphic concepts were commonly applied to various artificial software and robotic agents since at least the early 1990s, leading to important research directions and findings in AI and its applications spanning the late 20th and early 21st centuries. While I recognized the modeling and mental framework usefulness of such anthropomorphic approach to the properties and capabilities of artificial/engineered autonomous systems, I also argued about possible dangers and limitations of such approach; and in favor of cybernetics and systems science based approaches to the design, analysis and formal reasoning about autonomous AI-based systems [6]. Whether one approaches describing and reasoning about advanced AI systems from a social science and anthropomorphic standpoints or from a cybernetics and systems-science inspired perspectives, the promise of potential great benefits and well great destructiveness of such systems in general, and various types of fully autonomous unmanned vehicles in particular, was becoming increasingly apparent already shortly after the time of writing down those ideas as originally elucidates in [6], and then presenting them at research conferences in 2004 and

2005. In particular, upon further reading and thinking about how to best model and describe such AI systems, specifically focusing on intelligent robots and autonomous unmanned vehicles, I started having serious ethical concerns about the dangers of such systems in not-so-distant a future, especially in the context of contemporary and future war-waging. This has led me to begin considering the "moral agency" aspects related to the design and deployment of such autonomous AI-based systems in general, but especially in the context of warfare. While my own research interests have all along primarily been in peaceful, "civilian" applications of such technologies, ethical and legal dilemmas are still highly relevant even in those contexts, (obviously!) especially if and when "things go wrong", such as for example a self-driven car hitting and injuring or killing a pedestrian [7].

The important issue of *moral agency* of autonomously acting AI-based systems has become rather popular among the scholars in philosophy, ethics, humanities and social sciences; as well as among the legal scholars. Importantly, scholars in these non-technological disciplines have been exploring the matter of (moral) agency of AI systems combining the tools, techniques and conceptual frameworks from their own disciplines with those used by the technologists and (applied) scientists who actually research, design, implement, and/or formally analyze autonomous AI systems [8]. This cross-fertilization of ideas and modes originating in very different scholarly disciplines is rather encouraging. Certainly, a broad debate on the ethics and legal ramifications and challenges due to rapid rise of AI in general, and the use of AI in military applications and future armed conflicts in particular, should involve scholars and experts from a broad range of disciplines, from hard and applied sciences to government and industry leaders to legal scholars to scholars studying ethics and philosophy.

When it comes to the use of advanced AI and Robotics in warfare (as opposed to "civilian applications"), the very purpose of design and deployment of such systems is precisely to increase the "kill rate" of enemy's soldiers, while protecting one's own personnel (and presumably, protecting the civilian populations in the process). As extensive use of especially aerial drones in Iraq, Afghanistan and more recently Nagorno-Karabakh and Ukraine show, these AI-based weapons have indeed already become "very good" in increasing destruction and casualties to the enemy's military, but are nowhere as closely reliable insofar as protecting the civilians in or near combat zones from inadvertently (or sometimes even otherwise) becoming the so-called "collateral damage".

More recently, self-driven cars and other autonomous vehicles, their increasing use in "ordinary", civilian life, and the ethical and legal challenges that arise "when things go wrong" (such as, when an autonomous self-driven car causes a traffic accident; but also when a military drone, instead of hitting a terrorist or enemy base, ends up mistakenly targeting and killing a sizable group of civilians as has happened on multiple occasions in Afghanistan and elsewhere; see e.g. [9]) have attracted increasing attention of social scientists, legal experts, and scholars in humanities – specifically, those studying philosophy and ethics; see e.g. [10]. There are however several fundamental differences between the AI, robots and unmanned

vehicles causing harm in the civilian applications as compared with the use of those technologies in modern warfare. The first and most obvious one, is the express intent of human designers, as well as those deploying the technologies: in case of AI-driven, robotized warfare, the explicit intent and goal is to "maximize the harm" to the enemy (although, presumably, not to the civilians on the "enemy side" of the front-line). This explicit intent, combined with how devastating the AI and robotics based weapons already are (let alone how devastating they should be expected to become a few years, let alone a few decades, down the road!), makes the "ethical dilemma(s)" involved, in my view, in a sense easier to resolve: the authors of the Open Letter are undoubtedly fundamentally right, that the dangers to the survival of humanity let alone very realistic possibilities of massive scale destruction of life, will only grow as the AI-related science and technology continue rapidly progressing, and a broad, worldwide moratorium for the start, and ideally a complete ban, should take place as soon as possible! Some of the core problems related to this noble pursuit of a ban are shared with the nuclear race in the 20th century – from competing political and geo-strategic interests, distrust and fears of falling behind that we have seen in the context of nuclear armed race ("what if we stop developing *our* AI advanced weapons but others secretly continue developing *theirs*?"), lack of political will, and ultimately difficulties in enforcing compliance. With AI-based weapons, though, there is one novel core challenge. Namely, it has been, generally speaking, possible to monitor, observe and detect testings of nuclear weapons (as news article, research studies and policy analyses about such testings by "rogue regimes" such as that in North Korea show in recent years). However, monitoring and verifying that a nation, or another entity capable of developing and then using AI-based advanced weapons, is actually complying with a hypothetical ban and thus not doing any **R&D** or production of AI/Robotics based weapons systems will likely be nearly impossible!

We can already confidently assert that the Open Letter call from 2014 – 2015 has proven to be very timely, and hopefully it will lead to not just a healthy and diversified intellectual, scholarly and policy debates on the use of AI in modern and future warfare, but will also result in some measurable, globally supported action to limit the harms the mankind can inflict onto itself by using advanced AI-based weaponry. However, agreeing to broad international bans of advanced AI-based weapons has been, and will likely continue to be, very difficult if not elusive to accomplish. Even if and when such bans are adopted by the United Nations and other leading global organizations, practical enforcement of limits let alone total bans on AI-based weapons being researched, tested, manufactured and ultimately deployed in "real life", that is, in the future wars, will be extraordinarily hard to achieve in practice.

Summary

While exponential growth in AI research and technologies in recent decades holds a great deal of promise to improve almost all aspects of our lives from health care to energy generation and use to education, in the context of military applications it also holds great potential moral hazards and practical dangers to cause massive suffering and loss of human life. An increasing number of scholars, AI researchers, industry leaders and policy making has been calling on ban of AI-based weapons since at least 2014. However, achieving a globally adopted (and especially universally respected!) such ban will unfortunately be rather difficult. It remains of utmost importance that experts from different backgrounds and disciplines continue constructively engaging in discussing ethical, legal and practical ramifications of increasing use of AI-based autonomous systems in warfare, and seeking measures and solutions that will hopefully prevent our species from self-destructing via unhindered, short-sighted massive use of AI-based weapons in modern wars and potentially even letting autonomous AI systems wage future wars on "our" behalf. One may only hope that these debates will also eventually result in concrete, practical measures protecting the mankind from potential massive destruction resulting from unhindered use of advanced present and future AI-based weapons.

Acknowledgement

Conflict of Interest

References

1. The British Guardian article on the letter originally dating from 2014 and publicly presented at the leading AI conference (*International Joint Conference on Artificial Intelligence*, IJCAI) in Buenos Aires, Argentina back in 2015 can be found at https://www.theguardian.com/technology/2015/jul/27/musk-wozniak-hawking-ban-ai-autonomous-weapons (it was Guardian Online where I myself first read about this *open letter*)

2. The page where the Open Letter and its signatures are maintained: https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/

3. https://www.theguardian.com/politics/2015/apr/13/uk-opposes-international-ban-on-developing-killer-robots (April 2015)

4. (2021) Stuart Russell and Peter Norvig: "*Artificial Intelligence: A Modern Approach*", 4th US edition, Pearson.

5. Predrag T Tosic, Gul A Agha (2003) Understanding and Modeling Agent Autonomy in Dynamic Multi-Agent, Multi-Task Environments, in Proc. of the First European Workshop on Multi-Agent Systems (EUMAS'03), Oxford, England, United Kingdom.

6. Predrag T Tosic, Gul A Agha (2004) "Towards a hierarchical taxonomy of autonomous agents", Proc. IEEE Int'l Conf. on Systems, Man & Cybernetics SMC-04, The Hague, Netherlands.

7. Lauren Smiley (2020) "'I'm the Operator': The Aftermath of a Self-Driving Tragedy", *Wired* (online), March 2022. See also e.g. *CNN Business* article, "Uber self-driving car operator charged in pedestrian death" by Matt McFarland (September 2020) https://www.cnn.com/2020/09/18/cars/uber-vasquez-charged/index.html

8. Chris Graham (2011) "Autonomous Machines: Formalizing Moral Agency" (thesis paper): https://www.ilovephilosophy.com/viewtopic.php?t=177513

9. C Savage, E Schmitt, E Hill, C Koettl (2022) "Newly Declassified Video Shows U.S. Killing of 10 Civilians in Drone Strike", *New York Times* www.nytimes.com/2022/01/19/us/politics/afghanistan-drone-strike-video.html

10. Ryan M McManus, Abraham M Rutchick (2018) Autonomous Vehicles and the Attribution of Moral Responsibility", *Social Psychological and Personality Science*, SAGE Publishers: 1-8.