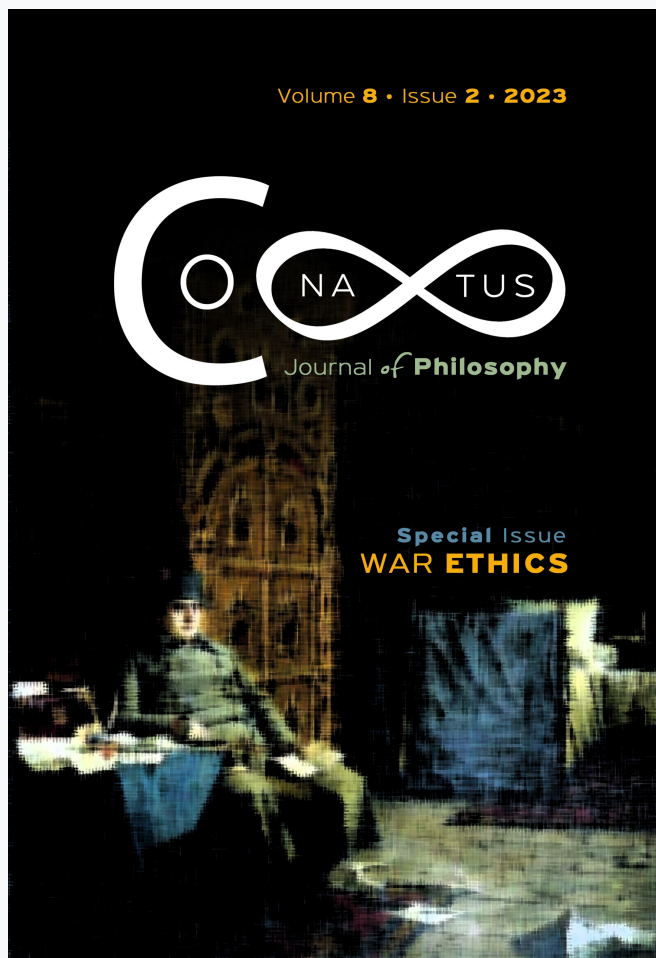


Conatus - Journal of Philosophy

Vol 8, No 2 (2023)

Conatus - Journal of Philosophy SI: War Ethics



Robotic Virtue, Military Ethics Education, and the Need for Proper Storytellers

Henrik Syse, Martin Cook

doi: [10.12681/cjp.35684](https://doi.org/10.12681/cjp.35684)

Copyright © 2023, Henrik Syse, Martin Cook



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

To cite this article:

Syse, H., & Cook, M. (2023). Robotic Virtue, Military Ethics Education, and the Need for Proper Storytellers. *Conatus - Journal of Philosophy*, 8(2), 667–680. <https://doi.org/10.12681/cjp.35684>

Robotic Virtue, Military Ethics Education, and the Need for Proper Storytellers

Henrik Syse

Peace Research Institute Oslo, Norway

E-mail address: syse@prio.org

ORCID iD: <https://orcid.org/0000-0001-5102-8079>

Martin L. Cook

U.S. Naval War College, United States (Emeritus)

E-mail address: martinleocook@gmail.com

ORCID iD: <https://orcid.org/0000-0002-9793-5958>

Abstract

The introduction of artificial intelligence (AI) challenges much of our traditional understanding of military ethics. What virtues and what sort of ethics education are needed as we move into an ever more AI-driven military reality? In this article we suggest and discuss key virtues that are needed, including the virtue of prudence and the accompanying virtue of good and proper storytelling. We also reflect on the ideal of “explainable AI,” and philosophize about the role of fear in helping us understand what is actually at stake in a military infused with AI-enabled and AI-driven weapons.

Keywords: *artificial intelligence; virtue; narratives; prudence; fear*

I. Introduction

In the opening chapter of Alasdair MacIntyre’s *After Virtue* from 1981, entitled “A Disquieting Suggestion,”¹ we find an imagined and indeed disquieting future portrayed. In that alternate timeline, the natural sciences have at some point been restricted and in essence

¹ Alasdair MacIntyre, *After Virtue* (Notre Dame, IN: University of Notre Dame Press, 1981), ch. 1.

purged, and their findings and results have been obscured and gradually forgotten. Later, however, they are recovered, but only in bits and pieces. The theoretical connections, the actual scientific experiments, and the underlying knowledge and insights that made the language of the natural sciences meaningful have been all but lost. Thus, the words are used, and scientific activity is seemingly being performed, yet the meaning and context of the endeavors make little sense. It is all essentially gibberish, but people do not know that. And, as MacIntyre notes with a hint of theoretical irony, analytic philosophy has no methods with which to reveal the problem, since all sentences and propositions do make sense within their own, internal framework.

For MacIntyre, this story parallels the way in which modern moral philosophy – in his view, dangerously imbued with emotivism and subjectivism – uses the language of virtue inherited from Aristotle but does not understand or appreciate it. We employ the terms, but we have lost the knowledge and the context that once gave them meaning. Hence, morality and its language become increasingly meaningless.

We do not have to agree wholeheartedly with MacIntyre’s diagnosis of our time to find the image a vivid one.

If one wishes for a more artistic impression of the same scenario, an episode of the TV series *The Twilight Zone* called “Wordplay” from 1985 will send chills down one’s spine.² A salesman, Bill Lowery, has accepted a job for a medical supply company, and to do his job, he must learn an advanced vocabulary totally unknown to him from before. In the course of this difficult process, he picks up some strange variations in the ways in which people around him speak, not just in their technical vocabulary, but in everyday language. First, he concludes that his colleagues’ jovial expression “Teaching old dogs new trumpets” must simply be part of the jargon. But when he is asked at lunchtime to join them for dinosaur, and the word “lunch” makes no sense to the others, he realizes something more dramatic is going on. Gradually, word by word, the language spoken by those around him becomes unintelligible, until he is totally shut out from the conversation of his fellows, lost in a maze of words that make no sense.

Our claim is that the rapid technological advances of our times create challenges along these lines. “You have forgotten to mute in Teams” is a sentence expressed daily now as we have mastered the intricacies of digital meetings, yet it made no sense before the Covid pandemic. While that is a perfectly innocent example, it arguably re-

² Wes Craven, dir., *The Twilight Zone*, season 1, episode 2a, “Wordplay” (Los Angeles: CBS Entertainment Productions, 1985).

minds us of the need to discuss – in the vein of MacIntyre – what happens to our language and understanding, including our *moral* language and understanding, as culture, technology, and language all change at breakneck speed.

We will venture two basic claims in the context of digitalization and so-called Artificial Intelligence (AI), especially as they relate to the ethics of armed conflict: Firstly, we need to study, rethink, and maybe even understand anew several of our traditional moral and intellectual virtues as we face an ever more digitalized world – and ever more digitalized conflict. What role can and do those virtues play as we increasingly work with and delegate tasks to intelligent, self-learning machines? And secondly, we may have to devise *new* virtues – or at least variants of the old ones – to fit with the challenges we face, not least in a military setting, from brain-computer interfaces employed by soldiers to virtual cyberwar and AI-enabled weapons. Are there virtues that we urgently need to formulate and emphasize?

This in turn is closely linked to our understanding and conception of military training and education, since a military force that has not been trained and educated to understand critically the world of artificial intelligence will also be using – and even be put in charge of – incredibly powerful machinery without grasping its implications and ramifications. If a crucial aim of military education is to strengthen those virtues that can guide soldiers to right action and right thinking, in accordance with the standards of just war tradition and international humanitarian law, we must assess how those virtues are to be understood and developed as we increasingly employ an AI-driven armed force.

As we discuss these very real challenges in the following, we will conclude by relating them to the fierce reality of fear and competition, elements that permeate our present-day world and must be taken into account as we discuss ethical ways of conducting armed conflict in light of complex, rapidly developing technologies.

II. Virtues

Virtues, understood as traits of character that are prerequisites for leading a good life, and which help us become good and well-functioning human beings, have traditionally been understood to be acquired in three ways: by learning, by training and habituation, and by being received as a divine gift. The first are primarily true of what we call the intellectual virtues, the second of the moral virtues, and the third of the theological virtues. Although we do not intend to downplay openness to the gifts of God, we will concentrate on the former two here.

What sorts of virtues do we need to operate and use advanced AI-guided military systems? Most obviously, in order to use, communicate, and act in close cooperation with a machine – indeed, to entrust one’s life and the success of one’s mission to that machine – we must have sufficient *understanding* of it so that we can trust it. We must understand its possibilities and limits, as well as the tasks to be carried out and the aims to be achieved. What virtues are required for this to be realized?

a. Courage

One of the main fears of critics of an increasingly automated and AI-driven military reality is the loss or denigration of the core value of honor and the accompanying virtue of courage.³ After all, machines have nothing to fear, but they also have nothing to be proud or ashamed of. Honor, conscience, the willingness to take risks, the courage required to put one’s life on the line: all of these may be lost at the altar of technology, or so it is claimed. Arguably, however, that is not true for the humans who develop, deploy, and operate such machines. *They* will still be afraid, feel shame, or experience honor.

In the *Laches*, the Platonic dialogue entirely devoted to the virtue of courage, one of the most promising definitions of that virtue to be arrived at is the following, slightly paraphrased: to be courageous is to know what one ought to fear. (The definition is suggested by the Athenian general Nicias at 194d-196c; we will leave aside here the problems they encounter with the definition). The question to us can be formulated as follows: Do we know what we ought to fear as we increasingly deploy AI-driven weapons?

In the military AI literature, the fear of losing meaningful human control, that is, of a slippage when it comes to the room for appropriate moral judgement and guidance, is often listed as *the* foremost challenge. Wrongly programmed machines may not come to be stopped before it is too late, or we will employ sophisticated algorithms without actually knowing and understanding how they will play out. This becomes an ever more real challenge as the complexity of digital machines as well as their interaction with other machines and algorithms rapidly increase. How do we confront these problems?

Formulating worst-case scenarios, defining points of no return, and having constant and readily available access to technical expertise that

³ See Valerie Morkevicius, “Tin Men: Cybernetics and the Importance of Soul,” *Journal of Military Ethics* 13, no. 1 (2015): 3-19, and M. Shane Riza, *Killing Without Heart: Limits on Robotic Warfare in an Age of Persistent Conflict* (Dulles, VA: Potomac, 2013).

can explain the functioning of the machines being used are all essential tasks and qualities that will enable the deployers and users of AI-driven military equipment to have and perform according to the virtue of courage. By courage, we mean here the moral and mental ability and willingness to venture into difficult yet important missions, even when using complex AI technology. Access to fine-tuned knowledge that tells us what to fear and what not to fear makes all the difference. Learning how to identify and tackle that fear and learning how to live with it in a way that leads to prudent decision-making – also about when and when not to use AI-driven systems – thus becomes a core part of military education overall, not least for the operators of such equipment.

b. Moderation

The virtue of moderation is, in the Platonic dialogue *Charmides*, also subject to several definitions (as was courage in the *Laches*), the most important and famous one being self-knowledge and self-control (formulated by Critias and Socrates together; see 164eff.). The truly moderate and balanced human being, possessing *sophrosyne* – soundness of mind, or moderation – is the one who honestly probes and knows his or her own strengths and weaknesses, possibilities, and limits. Socratic wisdom famously consists – most essentially – in knowing and appreciating *what one does not know*. In the context of AI, developing an increased awareness of the limits of not just the advanced machinery itself, which sometimes will seem more limitless than limited, but also of *our* limits in operating and understanding it, and the limits we can impose upon its use, lies at the heart of the virtue of moderation. It is arguably as relevant as ever, and also a core part of the training and education that must accompany AI-driven or AI-enabled weaponry. The feeling of being almost all-powerful when one utilizes such equipment, often accompanied by a sense of being physically safe, puts us in danger of obscuring the virtue of moderation and the key accompanying virtues of self-doubt and self-questioning.

c. Prudence

The virtue of prudence, understood as both an intellectual and a moral virtue – and not least as a *political* virtue – is certainly also made relevant and being challenged by the development of AI. Let us look at some aspects of its importance.

Firstly, the virtue of prudence describes a core quality needed for users of advanced technology such as AI in extreme contexts, namely, the

need to judge and appreciate the right levels of risk and uncertainty. It is a truism that engaging AI systems in a military setting should be done only when one can reasonably assume they will operate properly. But when is that the case? How can one make such a reasonable assumption?

One of the authors of this article has interviewed several commanders of naval vessels equipped with the AEGIS combat system. They express great reluctance about putting it in what they call autonomous mode and would indeed only consider it in the most contested environments. This exemplifies the sense of uncertainty surrounding any advanced system with great destructive power when it is not under human control. Of course, the actual level of uncertainty depends on the exact nature, complexity, and functioning of each individual system. For us, however, the crucial point is not to discuss concrete technologies, but to emphasize the importance of the ability of each commander of such systems to gauge the risk level rightly, based on experience, training, and knowledge of and familiarity with the systems used. This arguably encapsulates a core aspect of the virtue of prudence: the weighing of various forms of experience and evidence to make the right decision about launching weapons or operations with great destructive power.

On September 26, 1983, Soviet Lieutenant Colonel Stanislav Petrov disbelieved electronic warnings of an ongoing US nuclear attack on the Soviet Union, treating it instead as a false alarm. Had he done what he was supposed to do, namely, relay the alerts about a possible US nuclear attack to his superiors, it is likely that a Soviet nuclear counterattack would have commenced. Even though this incident took place at a time of significant tension between the United States and the Soviet Union, Petrov reasoned that the alarm (which, it turns out, was set off by sun being reflected off clouds) was false, not least because a US nuclear attack would have consisted of more rockets. Arguably, it was Petrov's prudence that saved the world from nuclear war.⁴ Had his hunch been wrong, the result would have been utter disaster for the Soviet Union. However, by employing his general knowledge of likely nuclear attacks, combined with what we might call his moral courage, he made a prudent decision, resting on his comprehension that even highly advanced electronic systems can be wrong. This is a clear instance of prudence in the handling of advanced technological systems.

Secondly, deep-seated knowledge of what AI systems are and can do will increasingly come to reside with specialists, due to the sheer

⁴ Alicia Sanders-Zakre, "The Man Who 'Saved the World' Dies at 77," *Arms Control Association*, <https://www.armscontrol.org/act/2017-10/news-briefs/man-saved-world-dies-77>.

complexity of AI. But prudence is needed well outside of those circles. Understanding what we are *actually* doing when we, for instance, develop new and ever more powerful forms of AI capabilities in the cyber domain, or decide to deploy and use such systems in battle theaters, is almost impossible if not accompanied by the right pedagogical tools for telling us what such a development entails. Such understanding must be based on know-how, experience, insight, and dialogue. Not without reason, several of the leading teachers of morality – from the Western philosophical and religious traditions, Plato and Jesus of Nazareth obviously come to mind – employ parables and analogies to penetrate to the deepest levels of moral understanding. This is prudence in practice, drawing on what we understand in order to venture into what we do not.

To understand fully and truly what a fast, complex, and ever-developing machine is doing, we thus need prudential, pedagogical work to explain the moral and practical ramifications of its existence and use. This is not done once and for all but must be refined constantly alongside the development of new machinery, programming, hardware, and software. True prudence depends on the ability to create narratives that clarify for all of us – soldiers, politicians, operators, and laymen alike – what is at stake. As Gregory Reichberg has pointed out in his masterful treatment of Thomas Aquinas and military prudence, war belongs within the realm of the constantly changing and the constantly uncertain. Great skill and soundness of mind are required to make decisions that are rightly ordered to the moral end of just warfare, namely, the common good.⁵ To hold together the immense complexity and uncertainty of military activities – not least in an age of advanced technology – with the defense of the common good and human dignity requires the kind of practical understanding that the virtue of prudence implies and facilitates. This virtue also requires the careful and wise construction of narratives that help us formulate how each action we perform belongs within a larger moral framework, and how the complex tools we employ may contribute to the societal and ethical ends of one's activity.

d. Storytelling and translation

This is where we suggest that we venture beyond the nomenclature of the virtues that are traditionally understood as important for the military sphere – and for military ethics – and propose an added virtue,

⁵ Gregory Reichberg, *Thomas Aquinas on War and Peace* (Cambridge: Cambridge University Press, 2017), 67-81.

built on what we just said about prudence: that of the skilled narrator, of the good and well-informed storyteller, who constantly, alongside the developers and the entrepreneurs – and in the military setting: the soldiers, commanders, and specialists – helps us translate the technology into understandable concepts and narratives and thereby assists us in grasping what we are doing, and where we may be going.

There is nothing new in this ideal or virtue *per se*; many such storytellers and translators exist, in academia as well as among writers of fiction. From Arthur C. Clarke and Isaac Asimov to Peter Singer and Stuart Russell, they have been among us for some time. But as a basis for political decision-making, for military ethics education, for operating manuals, and for military criminal tribunals in a world of AI, this virtue must be developed further.⁶

We do know this well from an analysis of our everyday lives. Things that affect us all the time, that are constantly deployed and used by us, are things we often do not understand, but which we nonetheless readily accept, even formally and explicitly, such as through repeatedly ticking “I accept” boxes. Let us provide a simple example: Recently, the daughter of one of the authors entered a restaurant with her father, and the latter asked her whether they hadn’t been there before, because it looked familiar – it’s just that we must have entered from the other side. She told him to look at Google, which rightly told him that he had been there almost exactly a year before. This in itself is trivial, but it does tell us a lot. When the author, at some point in the technically distant past, had ticked a box in Google allowing it to “use my location,” he had without much afterthought asked Google to keep track of his whereabouts. No one had told him a narrative, a parable, a story about what that actually meant, or translated his box-ticking into consequences he could readily understand. He had been confronted with many lines of legalese, of course, but very little to tell him what his “yes” signified in practice, and also – and importantly – no real sense of what the alternatives were.

We venture the claim that a primary job for philosophers, as well as for lawyers, social scientists, psychologists, and scholars of literature and religion in an AI-infused military, is to be, or to help identify, knowledgeable storytellers who, in close interaction with the engineers and developers of AI technology, attempt to tell us what the new technology implies, where we are heading, what we are doing, what will now be possible, and not least what the alternatives are. Maybe every

⁶ We have decided here more or less to conflate the “storyteller” and the “translator,” since the two functions overlap as well as complement each other in our context.

high-technology weapons manufacturer should be obliged to have a CSTO: a chief storytelling officer.

e. Virtue as a mean

In the tradition of Aristotle, a virtue is always a mean – a mean of excellence – between extremes. So, we need to think through what kind of a mean this just-mentioned quality of storytelling and translation might be.

We believe it is most essentially a mean between on the one hand the mastery of an impressive, yet purely technical language and insight, and on the other fantastical, fictional, often overly optimistic or downright scaremongering storytelling on the other. Both of these – the technical and the purely science-fictional – do have their place, and they are not vices *per se*. But as guides for political and moral prudence, they are sorely deficient, yet remarkably widespread.

In line with this, we would hold that incessant and heated debates between extreme optimists and equally extreme pessimists, even if well-intentioned, stand in danger of muddying the waters and hindering a balanced, morally alert conversation about AI.⁷ When the military AI debate becomes a battle between those who are certain this new technology will save the world and those who believe it will destroy it, we will stand in danger of losing sight of the real-life and truly momentous challenges we are facing.

III. Education and training

We believe that each of the virtues listed above should inform modern military education, and not least ethics education. We have entered a phase where soldiers at all levels will increasingly be expected to use and familiarize themselves with AI, in the form of (more or less) autonomous weapons and tools characterized by machine learning. Many of those tools will be extremely powerful means of warfighting and killing. To use them wisely and rightly will be our paramount task.

In order to accomplish that task, prudent pedagogy and truthful and accurate storytelling again come to the fore. Operating machinery of which one has no real understanding, whether as to its potential, limits, or consequences, is not only dangerous, but potentially destructive to the whole purpose of a common military enterprise. Understanding and

⁷ For a good illustration of these dangers, see Adam Lashinsky, “Marc Andreessen’s New Manifesto is a Self-serving Cry for Help,” *Washington Post*, October 19, 2023, <https://www.washingtonpost.com/opinions/2023/10/19/marc-andreessen-manifesto-silicon-valley/>.

guiding the most important means of fighting could end up becoming the domain of a few, while the bulk of operators and soldiers would be mere tools in their operation, in a way that eclipses the way in which a bomber pilot is a tool for the operation of a bomber. The bomber pilot will, after all, always have a basic, analogue understanding of the key means at his or her disposal. The user of powerful AI-guided weapons, who does not grasp their workings, their potential, or their limits, will be in a very different situation.

To avoid that scenario, the role of technology specialists as well as the “storytellers” and “translators” imagined above will be crucial, vis-à-vis both users (such as soldiers and other operators), decisionmakers in the war theater, and political decisionmakers. Furthermore, and as a consequence of this, education in military ethics for soldiers must aim to strengthen both the intellectual and moral understanding of what these weapons and systems most deeply represent. That also entails a thorough and broad dialogue between specialists, decisionmakers, and users at all levels.

This does not mean that all levels of a military force will need the same detailed understanding of the tools used. For the enlisted and lower-level officers, basic training so that one knows how to operate the system is required, and also, if there are known dangers in trusting the system, additional training to recognize those limit situations and exercise appropriate caution. Examples of military personnel *not* trusting their systems to good effect are plentiful – such as learning to recognize when a gun is about to misfire, or a flight compass is giving faulty readings. Hence, doing whatever possible to train for when to trust and when to override is critical. At the same time, in very time-critical applications, we do not want to train or educate for excessive caution either, which could obviously be lethal. As with so much of virtuous action, the ideal will consist in a mean between caution and efficiency, between healthy skepticism and well-placed trust. But we emphasize again that we have no illusion that all users of advanced AI-enabled weaponry will or can be fully educated in the technological and ethical aspects of using such weaponry. The basic training they do receive, however, must be of a kind that heeds the lessons of those who truly understand the technology, and makes the user aware of the limits of each particular technology, weapon, and system.

IV. Explainable AI?

Our vision here of good storytelling, and accompanying training and education, as a key part of making AI practically and ethically employable echoes the ideal of “explainable AI.” The underlying idea of this

phrase is that the functioning of the tools we use should be possible to explain in a way that makes it doable to understand, predict, and *post facto* re-trace what the tools actually do.

In its strict sense, it is clearly not a realistic goal. Even everyday digital and digital-enhanced instruments we use and on which we are dependent, from cars to personal computers, are such that it is not possible to explain all their elements to the lay user. Indeed, explainability in the full sense, pertaining to the user, should not be a constraint on AI development. However, we should be able to expect predictable input-output processes, even when – to many or even most users – the contents of the algorithms will remain a black box. In other words, rigorous testing should ensure that we can have high confidence that a given set of inputs will give us the output we can rely on. Proper education, technical as well as ethical, will have as its main task to make commanders awake to the possibility that this might not always be the case, and that outputs must be monitored to ensure that the AI systems are truly working as they should.

For example, using AI for target nomination on the battlefield is clearly a crucial application for such systems. That is, the AI system can identify a tank, an artillery piece, or a command post. Those identifications will not necessarily be extremely time critical, so presumably human targeteers can examine those nominations and agree or disagree. In more time-critical situations, such as defending a ship against high-speed or hypersonic incoming missiles, the human ability to engage fully in every step of the OODA (observe, orient, decide, and act) loop will be limited, and so systems such as the naval AEGIS Combat System in autonomous mode will have to be relied on for lack of a good alternative. However, even here observing input-output reliability will be absolutely crucial, combined with the necessary assurance that the commander deciding to use that system in autonomous mode knows about the system's workings and safety, and can both monitor and *ex post* explain its use.

V. Concluding reflections

We started out with stories from Alasdair MacIntyre and *The Twilight Zone*, both of them simultaneously insightful, scary, and entertaining. If – ten or twenty or fifty years from now – we still speak of courage and moderation, of proportionality, authority, and intentions, yet our language has little to do with the actual world we inhabit, or the actual moral and political challenges confronting us, we may indeed be in trouble. Equally troublesome would be a world where we use these

traditional words from the history of moral philosophy and just war tradition, yet do not understand what they mean, because we have been overwhelmed by a development that has simply scrambled their meaning.

For this to happen, we do not need AI, of course. Moral degradation, extreme nationalism, or authoritarian ideologies have proven themselves worthy enablers of such dissembling many times before, quite apart from technological prowess. To hear Foreign Minister Sergey Lavrov or President Vladimir Putin of Russia repeatedly tell us that in the invasion of Ukraine, no civilians are being targeted, and that the “Nazi” regime in Kyiv needs to be driven out, is a chilling reminder. The words have no truthful meaning, but they function within their own closed world, what Robert Musil and Eric Voegelin famously called a Second Reality.⁸ The world of AI could become such a world. That is what we must avoid.

Therefore, we need a constantly evolving, deep-seated conversation about morality, and we need the human virtues of courage, moderation, prudence, and justice,⁹ accompanied by good and informed storytelling – encompassing also interpretation and translation – to guide us. Most likely, the machines themselves will never possess those virtues in any real, conscious sense; that belongs to human beings. It is all the more important that their human developers, enablers, and operators possess them. To possess them, we need to understand the technologically extremely advanced world into which we are moving. Proper, balanced, moderate, and knowledgeable narrators are key to that endeavor.

This vision of storytelling and appropriate virtues must not obscure a realistic assessment of where we find ourselves, namely, in a *de facto* arms race. Several participants in that race may have little interest in constructing a common ethical narrative about AI. The Chinese and (albeit to a less technologically advanced degree) the Russians seem to be plowing ahead with developing these technologies with less worries about constraints than the West currently has. By saying this, we are not claiming that a lacking attention to ethics is a feature of one side only, or that there is no interest in ethical or safety-related constraints

⁸ T. John Jamieson, “Robert Musil and Eric Voegelin: Literature and Spiritual Pathology,” *VoegelinView*, May 10, 2012, <https://voegelinview.com/robert-musil-and-spiritual-pathology-pt-1/>.

⁹ We have not dealt with justice in any detail in this article. We do mention it here, though, since it is clearly crucial to any theory of just war, and also since it is considered in the philosophical tradition to be one of the cardinal virtues, alongside moderation, courage, and prudence.

in a country such as China. The point is that in an arms race, the danger of overlooking key ethical or human-rights-related issues, in the name of the nation's (or the party's) interests, is an ever-present danger.

This situation also reminds us, however, that we have a virtue – albeit often portrayed as a vice – that can be our friend, namely, fear, which we also touched on above in our discussion of proper courage. We are dealing with potentially very dangerous technologies, and all rational actors can see that this is the case. For that reason, like with nuclear disarmament, fear may be the best common motive for action. Rational fear, mutually shared, provides a basis for shared interests, which might, in turn, provide a framework for agreed restraint and regulation.

Rational fear requires rational, shared narratives, which can be shared between technologists as well as politicians. There are those who deeply distrust a rules-based order grounded in international law and want to replace it with a much more multipolar world. That could, we fear, lead us to a Hobbes-like state of nature, with powerful technologies set to fuel the fires of international distrust and conflict. The fear of such a world should be a guiding light for us all.

We must not give up hope for a new San Francisco – the founding city of the United Nations – for the 21st century, in which new rules are agreed for effective transnational institutions with viable enforcement mechanisms, not least when it comes to AI. While that may seem unlikely at the current moment, it seems clear that its basis and best friend would be fear of the consequences of what happens if we do not get there, and if wars get completely out of hand and more and more parties do not even pretend to follow the rules (or even program their weapons not to do so). But for that fear to be rationally and factually formulated we, again, need strong narratives, good and rational storytellers, brilliant translators (literally and figuratively), and educated decisionmakers who truly understand the perils of the situation.

Acknowledgements

Work on this article has been funded by the Research Council of Norway via a project called “Warring with Machines: Military Applications of AI and the Relevance of Virtue Ethics.” A draft version of the article was presented at the McCain Conference on Military Ethics at the United States Naval Academy in 2022. We are most grateful to the Research Council of Norway, to our colleagues in the Warring with Machines project, and to Dr. Edward Barrett, main convenor of the McCain Conference on Military Ethics.

References

Craven, Wes, dir. *The Twilight Zone*. Season 1, episode 2a, "Wordplay." Los Angeles: CBS Entertainment Productions, 1985.

Jamieson, John T. "Robert Musil and Eric Voegelin: Literature and Spiritual Pathology." *VoegelinView*, May 10, 2012. <https://voegelinview.com/robert-musil-and-spiritual-pathology-pt-1/>.

Lashinsky, Adam. "Marc Andreessen's New Manifesto is a Self-serving Cry for Help." *Washington Post*, October 19, 2023. <https://www.washingtonpost.com/opinions/2023/10/19/marc-andreessen-manifesto-silicon-valley/>.

MacIntyre, Alasdair. *After Virtue*. Notre Dame, IN: University of Notre Dame Press, 1981.

Morkevicius, Valerie. "Tin Men: Cybernetics and the Importance of Soul." *Journal of Military Ethics* 13, no. 1 (2014): 3-19.

Reichberg, Gregory. *Thomas Aquinas on War and Peace*. Cambridge: Cambridge University Press, 2017.

Riza, Shane M. *Killing Without Heart: Limits on Robotic Warfare in an Age of Persistent Conflict*. Dulles, VA: Potomac, 2013.

Sanders-Zakre, Alicia. "The Man Who 'Saved the World' Dies at 77." *Arms Control Association*, 2017. <https://www.armscontrol.org/act/2017-10/news-briefs/man-saved-world-dies-77>.