



AFRL-AFOSR-VA-TR-2023-0316

**THE WARFIGHTER'S TOLERANCE FOR AUTONOMY AND ITS
IMPORTANCE IN
STRATEGY & SYSTEMS DEVELOPMENT**

**Jai Galliot
University of New South Wales-NEW NCAGE Code
HIGH STREET
KENSINGTON, NSW, ,
AU**

**03/22/2023
Final Technical Report**

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20230322		2. REPORT TYPE Final		3. DATES COVERED	
				START DATE 20180701	END DATE 20211031
4. TITLE AND SUBTITLE THE WARFIGHTER'S TOLERANCE FOR AUTONOMY AND ITS IMPORTANCE IN STRATEGY & SYSTEMS DEVELOPMENT					
5a. CONTRACT NUMBER		5b. GRANT NUMBER FA9550-18-1-0181		5c. PROGRAM ELEMENT NUMBER 61102F	
5d. PROJECT NUMBER		5e. TASK NUMBER		5f. WORK UNIT NUMBER	
6. AUTHOR(S) Jai Galliot					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of New South Wales-NEW NCAGE Code HIGH STREET KENSINGTON, NSW AU				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/AFOSR RTA2	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-AFOSR-VA-TR-2023-0316	
12. DISTRIBUTION/AVAILABILITY STATEMENT A Distribution Unlimited: PB Public Release					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence. The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of "tolerance", i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier's (or his unit's) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier's tolerance for technology can be preserved only if automation is respectful of the "power process", i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals. The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:				17. LIMITATION OF ABSTRACT	
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU		18. NUMBER OF PAGES 13
19a. NAME OF RESPONSIBLE PERSON LAURA STECKMAN				19b. PHONE NUMBER (Include area code) 426-7556	

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20230322		2. REPORT TYPE Final		3. DATES COVERED	
				START DATE 20180701	END DATE 20211031
4. TITLE AND SUBTITLE THE WARFIGHTER'S TOLERANCE FOR AUTONOMY AND ITS IMPORTANCE IN STRATEGY & SYSTEMS DEVELOPMENT					
5a. CONTRACT NUMBER		5b. GRANT NUMBER FA9550-18-1-0181		5c. PROGRAM ELEMENT NUMBER 61102F	
5d. PROJECT NUMBER		5e. TASK NUMBER		5f. WORK UNIT NUMBER	
6. AUTHOR(S) Jai Galliot					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of New South Wales-NEW NCAGE Code HIGH STREET KENSINGTON, NSW AU				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/AFOSR RTA2	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-AFOSR-VA-TR-2023-0316	
12. DISTRIBUTION/AVAILABILITY STATEMENT A Distribution Unlimited: PB Public Release					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence. The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of "tolerance", i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier's (or his unit's) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier's tolerance for technology can be preserved only if automation is respectful of the "power process", i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals. The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:				17. LIMITATION OF ABSTRACT	
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU		18. NUMBER OF PAGES 13
19a. NAME OF RESPONSIBLE PERSON LAURA STECKMAN				19b. PHONE NUMBER (Include area code) 426-7556	

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:

- 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliott

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliott

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliott

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliott, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliott (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliott J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:

- 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliott

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliott

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliott

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliott, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliott (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliott J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliott

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliott

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliott

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliott, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliott (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliott J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliott

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliott

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliott

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliott, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliott (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliott J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliott

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliott

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliott

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliott, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliott (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliott J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliot

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliot

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliot

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliot, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliot (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliot J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliot

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliot

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliot

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliot, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliot (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliot J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliott

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliott

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliott

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliott, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliott (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliott J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliott

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliott

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliott

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliott, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliott (eds.), *Big data and democracy* (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliott J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliot

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliot

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliot

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliot, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliot (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliot J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliott

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliott

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliott

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliott, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliott (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliott J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliott, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliot

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliot

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliot

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliot, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliot (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliot J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.

Technical Report - FA9550-18-1-0181

You will be asked to complete the following questions as part of the report documentation section:

- Distribution Statement –Please verify that the report you are about to upload is cleared for public release through your organization’s processes and procedures.
- Report Abstract
- Report Document - Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 100MB.

Regarding the content of the report, we require the following sections in your PDF technical report upload.

Abstract

The project has systematically explored the attitudes, beliefs, and aptitudes of military personnel in relation to autonomous military technologies, with a specific focus on autonomous weapon systems. In the first place, this investigation intended to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing military human workforces with autonomous military technologies; in the second place, the project also aimed to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities and risk mitigation strategies. The core problem investigated by this project is that humans tend to reject autonomous agents when they have either a lack or an excess of trust towards such agents or again when they have a poor familiarity with the relevant systems, due to the inability to establish with them reciprocal coordination and understanding. This problem is perceived as serious especially in the context of hybrid military operations in which human warfighters must entirely rely on the information shared by/with synthetic agents and artificial intelligence.

The guiding assumption of our interdisciplinary work in the legal and social sciences is that, while until now the human-machine relationship has been prevalently addressed as a technological challenge, it can in fact only be properly understood in a much broader socio-philosophical context. This context is centrally defined by the notion of “tolerance”, i.e. the capacity of soldiers to endure and withstand their subjection to technology despite the fact that automation impinges upon their personal autonomy and possibly impacts the soldier’s (or his unit’s) psychological wellbeing. The project initial hypothesis, which has been confirmed through the analyses developed by the participants, is that the soldier’s tolerance for technology can be preserved only if automation is respectful of the “power process”, i.e. the socio-psychological dynamics in which healthy human beings freely exert their individual decision-making capability, responsibility, and agency while effortfully striving to achieving their personal goals.

The project has prepared the theoretical background and the methodological tools necessary to assess the psychological downsides, the legal and moral constraints, and the socio-cultural requirements for the adoption of autonomous technologies, evaluating the toll that artificial agents put on the armed forces development capability. The research conducted by our group

has examined opinions and beliefs related to personnel recruitment, reward and recognition, promotion, and skills divide, as well as risks associated with psychological and socio-cultural forms of displacement, including alienation, traumas, undesirable anthropomorphic projections, and more intangible downsides like moral deskilling and moral injury. The researchers have drawn several conclusions concerning the functionality of the systems and the possible responses of the human users, producing models and indications that might be concretely applicable to the design of future autonomous artificial agents and effectively useful to identify desirable policies and best practices to be implemented.

Accomplishments

The information provided in this section allows AFOSR to assess whether satisfactory progress has been made during this reporting period. The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

- Research Objectives: Please list the main research objectives of this project
 - List the major goals of the project including the scientific or technological objectives of this effort. Describe the proposed technical approach to obtain those goals. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The project aimed to fulfil two main goals: first, to delineate areas of risk and ethical uncertainty associated with the prospect of replacing or complementing human military workforces with autonomous military technologies; second, to weigh such risks and uncertainties against the inherent tactical and operational advantages represented by autonomous technologies, identifying viable opportunities of hybridization and risk mitigation strategies. The project aimed to fulfill these two goals by achieving seven major objectives. The first research objective consisted in “determining how the power process ought to be conceptualized for the purpose of examining relevant interactions between humans, automation science and relevant engineering/systems”. This objective was entirely met through the publication of Article #1 (Galliot 2018). The manuscript introduced the notion of ‘power process’ as a theoretical construct in the social sciences and implemented it systematically to critically review the problems that are most likely to affect the military by reason of the progressive introduction of autonomous technologies.

The second research objective required to “evaluate the extent to which autonomous systems aid fulfilment of the power process and thus promote tolerance of autonomy”. This objective was fully achieved through Article #2 (Jevglevskaja & Galliot 2020), which collected illuminating information about the attitudes of air force personnel toward unmanned aerial vehicles and examined their implications for the future of military aviation.

The third research objective asked the researchers to “evaluate the extent to which autonomous systems disrupt the power process and analyze the implications any disruption of the power process has on soldiers’ socio-psychological welfare, their units and the relevant operating environment/s”. This objective was initially explored by Article

#3 (McFarland & Galliot, not yet published), dedicated to skill divide, and Article #4 (McFarland, Galliot & Cappuccio 2020), dedicated to the big data paradox. These articles examined some modalities of workforce disruption that are largely known in the civilian context and draw some conclusions concerning the implications in the Defence domain. Subsequently, the objective was fully accomplished thanks to Article #5 (Guha & Galliot, not yet published) dedicated to moral deskilling, and Article #6 and Article #6bis (Cappuccio, Galliot, & Sandoval 2021 and 2021bis), dedicated to anthropomorphism and empathy in soldier-autonomy teams, two important unwanted psychological consequences arising in the context of hybrid military operations involving human warfighter and artificial agents. Together, these five research papers explore the psychological and sociological impact exerted by autonomous technologies on military personnel, assessing the real extent of their disruptive aspects.

The fourth research objective consisted in “assessing how soldiers are likely to adjust beyond simply disengaging from technology and consider how any negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”. Four different forms of psychological and moral adjustment are envisioned by Article #7 Rupka, Article #8 (Cappuccio, Galliot & Alnajjar 2022 forthcoming), Article #9 (Guha & Galliot, not yet published), and Article #10 (Jevglevskaja & Liivoja 2021). These articles respectively focus on military reward, moral injury, revolt & sabotage, and military recruitment to examine how the classical patterns of behaviors, beliefs, and values need being modulated to make room for autonomous systems and to update the protocols and practices governing the work of human beings accordingly.

The fifth research objective requested the research group to “build understanding of how autonomous technologies can be better designed, developed and engineered to limit disruption of the power process and better account for their practical and normative implications, thus building tolerance and trust in autonomous systems”. This research objective was achieved by Article #11 (McFarland & Galliot 2021), Article #12 (Soltanzadeh, Galliot & Jevglevskaja 2020), Article #13 (Cappuccio, Galliot, Eyssel, Lanteri, under review), and Article #14 (Cappuccio, Galliot, Eyssel & Lanteri 2021).

Dissemination was the sixth objective of this research project. It was specifically pursued by the deliverables numbered from #15 to #24, which implemented different forms of dissemination through a variety of media and communication methods, including a dissemination plan, a project website, campaigns on social media, electronic newsletters, and conference presentations/workshops. Eventually, the conference presentations and the final workshop specifically conceived by the PI to meet this objective (deliverables #24-25) were not executed out due to the disruptions caused by the Covid-19 pandemics and the extensive periods of lockdown associated with it.

- Please provide details of accomplishments during this reporting period.
- For this reporting period describe:
 - 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative);

The project's major activities consisted in the production of original research output, namely a series of research papers intended to be published in peer-reviewed journals

or academic books, plus one edited monograph (a collective volume) and one mini course curriculum. As detailed above, the project has fulfilled in a satisfactory manner all the main research objectives and it has provided the researchers involved with the concepts and tools necessary to independently continue and expand this investigation through their future projects. The main objectives include: “understanding how autonomous technologies can be better designed to limit disruption of the power process”; understanding “how negative implications might be averted or mitigated via the adjustment of relevant policy or strategic settings”; understanding “how any negative implications might be averted via the adjustment of relevant policy”; and “evaluate the extent to which autonomous systems disrupt the power process or, when appropriately designed, aid fulfilment of the power process and thus promote tolerance of autonomy”. The project has both met and exceeded the general expectations concerning the assessment of tolerance and the analysis of the psychological and socio-cultural dynamics that accompany the disruption of the ‘power process’ in the context of human-autonomy interactions. The project has successfully tested and implemented a general approach based on the interdisciplinary integration of different methodological perspectives such as critical political discourse, rigorous analysis of philosophical argumentation and legal reasoning, as well as qualitative and quantitative methods from cognitive science and sociological research. The application of this methodology to the project’s main themes has ultimately confirmed the essence of some ancient philosophical concerns about technological alienation, which essentially consists in the failure of the personal and interpersonal processes that regulate agency, independency of thought and action, and personal purpose and motivation due to technology that display human systems of values, relationships, or public roles and functions. The research conducted by the team members who applied this methodology has produced the following findings:

“Tolerance” for autonomous systems, the notion introduced by Dr. Jai Galliot to describe the soldier’s endurance to subjection to technological systems has proven to be a nitidly defined, versatile, reliable, and methodologically robust social sciences construct. Also, it is consistent and compatible with other, similar constructs commonly used in human-machine interaction research, such as Trust and Acceptance, while being fundamentally irreducible to them. As a metric inspired by the strict methodological and procedural standards of social science research (and, to a minor extent, by the research in human factors & ergonomics), Tolerance can be effectively implemented in the design and customization of human-machine interaction applications, where it can be used to measure with specifically devised scales the soldier’s endurance or insusceptibility to the disruptions of the power process caused by the sustained exposure to autonomous artificial agents. Distinguishing Tolerance from Trust and Acceptance proves indispensable to specifically capture the relevant forms of hostility and rejection behaviors toward technology, i.e. the negative responses that are distinctively motivated by deeply intrenched biases and prejudices against intelligent machines. These represent the distinctive feature of the unsympathetic, antagonistic and skeptical reactions towards the autonomous systems that have been recently adopted by the armed forces. The notion of Tolerance is particularly useful when combined with the concept of ‘power process’, which has demonstrated its pervasive applicability and usefulness in decoding various elements of the military profession and various aspects of the decision-making and governance processes in the space of security and defence, including recruitment, training, reward and recognition, promotion, enforcement of codes of ethics, enforcement of orders via authority, chain of command,

allotment of responsibilities. This analysis has revealed how all human agents fundamentally aspire to stabilize or expand their opportunities to express their potential for autonomy and self-determination as these are key components of their mental health and personal well-being. The collective preservation of the power process is a crucial concern in the context of military life, a professional context that is obviously characterized by martial discipline, adherence to public values and recognition of collective identities, controlled aggressive impulses, and the utmost respect for strongly hierarchical decision-making processes. This context presupposes a system of robust and reliable interpersonal relations that functions more efficiently when it promotes loyalty and cohesiveness without systematically imposing the suppression of natural impulses, individual goals, and personal drives. In this sense, the possibility that autonomous technologies interfere with the power process is ultimately negative because it exacerbate risks and pre-existing instabilities; such risks have always been associated with the psychological inhibitions that contribute to shape the individual autonomy of humans, being co-extensive with conflicts and tensions that were already known within the military culture before the age of autonomous technologies and AI. This means that the potential alienation and the risks of rejection and even revolt associated with the technologization of lifestyles and the automation of work processes and interpersonal relations echoes and exposes in a more vivid manner the dramatic sense of alienation that some individuals associate with a military culture that they apparently are unwilling to embrace but that they are not ready to change either. That is why defusing the risks arising from the most invasive applications of autonomous technologies in the workplace can be done using risk management approaches and interventions that have already been effective to monitor and defuse the risks of revolt and insubordination in the military domain.

Ultimately, our project has indicated that some concepts, models, and methods are more effective than others in identifying the potential disruptions of the power process associated with low tolerance to technology. Also, the concepts, models, and methods identified by our research seem more promising in relation to the need to mitigate the negative effects caused by these potential disruptions, as they can support specifically trained operators in their task to help personnel reconstruct a healthy and functional relationship of trust and acceptance with autonomous artificial agents, defusing their negative feelings and their beliefs that they are becoming passive instruments at the service of technology.

4) other achievements. Include a discussion of stated goals not met.

The main goals, related to the production of original research contents, have been met and exceeded: the group was eventually able to deliver all the 14 Articles originally envisioned by Dr. Jai Galliot as part of the action plan (Deliverables #D1-#D14) + one edited monograph (Deliverable #D19) + all the four additional articles envisioned as part of the project extension were effectively completed. In short, the research outputs produced by the team members over the period of the grant include 10 manuscripts (i.e., 9 articles + one edited monograph) published or accepted for publication in well-respected international journals or collective volumes and 10 articles (which are currently undergoing peer-review or waiting to be submitted or re-submitted). It must be noted that, while publishing 10 deliverables is a positive outcome in itself, the unexpected prolongation of the average review time due to the pandemics has caused some delays to our publication plans, which is why a number of the remaining deliverables are still undergoing peer-review

process.

Article #1 PUBLISHED

Jai Galliot

The soldier's tolerance for autonomous systems

[Paladyn, Journal of Behavioral Robotics](#)

2018; 9:124–136

<https://doi.org/10.1515/pjbr-2018-0008>

Article #2 PUBLISHED

Dr. Natalia Jevglevskaja & Dr. Jai Galliot

Airmen and Unmanned Aerial Vehicles: The Danger of Generalization

Wild Blue Yonder, April 13, 2020

NO DOI, but there is a link:

<https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2144225/airmen-and-unmanned-aerial-vehicles-the-danger-of-generalization/>

Article #3 NOT PUBLISHED YET

Tim McFarland & Jai Galliot

Robotics, Artificial Intelligence and the Increasing Skills Divide in the Armed Forces

Article #4 PUBLISHED

Tim McFarland, Jai Galliot, Massimiliano Cappuccio (2020)

The Big Data Paradox and its Importance to Strategy and Systems Development

In Kevin Macnish, Jai Galliot (eds.), Big data and democracy (2020)

Edinburgh University Press

ISBN: [9781474463522](#)

<https://espace.library.uq.edu.au/view/UQ:b39e037>

Article #5 NOT PUBLISHED YET

Guha M, Galliot J, 'Autonomous Systems and Moral De-Skilling: Beyond Good and Evil in the Emergent Battlespaces of the 21st Century'

Article #6 PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Mapping Meaning and Purpose in Human-Robot Teams: Anthropomorphic Agents in Military Operations

The Philosophical Journal of Conflict and Violence Vol. V, Issue 1/2021

DOI: 10.22618/TP.PJCV.20215.1.139005

<https://trivent-publishing.eu/img/cms/5-%20Massimiliano%20L-%20Cappuccio%20et%20al-.pdf>

Article #6 bis PUBLISHED

Massimiliano Cappuccio, Jai Galliot, Eduardo Benitez Sandoval (2021)

Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams.

International Journal of Social Robotics. 1-14

<https://doi.org/10.1007/s12369-021-00755-z>

Article #7 NOT PUBLISHED YET

Sean Rupka, Jai Galliott
Rethinking Military Reward and Recognition Systems in the Age of Autonomy

Article #8 ACCEPTED FOR PUBLICATION (WAITING DOI)

Massimiliano Cappuccio, Jai Galliott, Fady Alnajjar (forthcoming 2022)

A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury

Journal of Military Ethics

Article #9 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Rage Against the Machine. Military Technology, Revolt and Sabotage: Internal Threats Originating from Discontent

Article #10 PUBLISHED

Natalia Jevglevskaja and Rain Liivoja, The Better Instincts of Humanity: Humanitarian Arguments in Defense of International Arms Control In: Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare. Edited by: Jai Galliott, Duncan MacIntosh and Jens David Ohlin (eds. 2021).

Article #11 PUBLISHED

Tim McFarland & Jai Galliott (2021)

Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots

In Jai Galliott, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)

[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)

Oxford University Press

DOI:10.1093/oso/9780197546048.003.0004

Article #12 PUBLISHED

Sadjad Soltanzadeh, Jai Galliott, Natalia Jevglevskaja (2020)

Customizable ethics settings for building resilience and narrowing the responsibility gap:

Case studies in the socio-ethical engineering of autonomous systems

Science and Engineering Ethics, 26(5),

Springer: Netherlands

pp. 2693-2708

DOI: [10.1007/s11948-020-00221-5](#)

Article #13 NOT PUBLISHED YET

Manabrata Guha & Jai Galliott

Value Sensitive Design for Autonomous Systems

Article #14 NOT PUBLISHED YET

Massimiliano Cappuccio, Jai Galliott, Friederike Eyssel, Alessandro Lanteri

Autonomous Technologies and the Monitoring of Acceptance, Trust, and Tolerance

Additional Article #1 NOT PUBLISHED YET

Jai Galliott, Massimiliano Cappuccio, Austin Wyatt

Taming the Killer Robot: Toward a Set of Ethical Principles for Military AI

Additional Article #2 NOT PUBLISHED YET
Manabrata Guha

Additional Article #3 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Additional Article #4 NOT PUBLISHED YET
Massimiliano Cappuccio, Jai Galliot, Alessandro Lanteri

Deliverable #19 Monograph PUBLISHED
Jai Galliot, Duncan MacIntosh, and Jens David Ohlin (eds. 2020)
[Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare](#)
Oxford University Press
Print ISBN-13: 9780197546048

Some of the other project objectives, especially those instrumental to reach the “Dissemination” and “Professional Development” Objectives (Deliverables #15-24) were only partially met, as the Covid-19 pandemics diminished our group’s potential to effectively communicate the results of our research in the broader society and reach out to the scientific community, as the extended periods of isolation and lockdown concretely undermined our capability to present our research papers at local and international conferences, host our own scientific workshops, and organize networking events (as per Deliverable #24).

- How were the results disseminated to communities of interest? If there is nothing significant to report during this reporting period, state “Nothing to Report.”
- Describe how the results were disseminated to communities of interest for this reporting period. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated through the publication of an edited Monograph for a major international publisher, which gave great academic visibility to our project (Deliverable #19, co-edited by Dr. Jai Galliot with Prof. Duncan MacIntosh). Also, dissemination relied extensively on social media activities (coordinated by Dr. Sadjad Soltanzadeh as per Deliverable #16) and official communication occurred through the group’s official website and professional blog (coordinated by Dr. Tim McFarland, as per Deliverable #17), which included, inter alia, abstracts and executive summaries of the Articles, and finally through the electronic newsletters (prepared by Dr. Massimiliano Cappuccio and circulated by Dr. Jai Galliot, as per Deliverable #23). Also, some events organized and held by our group during the period of the Minerva grant were instrumental to allow the group members to present their research and engage with international scholars representing both the academia and the Armed Forces: first, the Ethics of Law of Trusted Autonomous Systems conference, which was held on 22-24 May 2019, Lady Margaret Hall, Oxford University; second, the TEMPER workshop on Training Enhancement and Military Performance, held at ADFA on 13-14 February 2020.

- (Interim Reports Only) What do you plan to do during the next reporting period to accomplish the goals and objectives? If there are no changes to the agency-approved application or plan for this project or if this is the final report, state “Nothing to Report.”
- Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report (this is the project's final report).

Impacts

This component is used to describe ways in which the work, findings, and specific products of the project have had an impact during this reporting period. Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project. You can report on the following impact categories, but you are not required to report on all categories. Please only report in the categories that are relevant to your project:

Development of the principal discipline(s) of the project

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that a lay audience can understand (e.g., Scientific American style).

The publication of our research papers has attracted attention to the attitudes and beliefs of military personnel toward autonomy, raising a number of ethical, psychological and cultural issues that institutional stakeholders will have to consider for the future design of new technologies and policies and for the implementation of innovative practices and processes.

Other disciplines:

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Describe the impact in this reporting period on the development of human resources

For example, how has the project provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; provided scholarships; provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?

The research and writing skills of all the team members have improved during the time spent in our group and significantly developed because of the tasks they had to carry out individually or collectively under the Minerva project. The professional experience acquired in the context of this research project has prepared most of them to move to the next stages of their careers: Dr. McFarland, Dr. Baggiarini, Dr. Jevglevskaja, Dr. Cappuccio and Dr Guha have been promoted before or immediately after completing their work for this

project. Mr. Sean Rupka is currently benefitting from this experience to write his PhD dissertation.

Describe the impact on teaching and educational experiences

For example, has the project: Developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?

The project has solicited an intense reflection, among the team members, about key themes and problems in technology ethics. This reflection was specifically articulated to respond to an objective demand for training and skill development in these areas, a demand coming more and more prominently from professionals working in the domains of security of Defence. This reflection has informed the contents and the goals of the online mini course specifically developed by Dr Guha (Deliverable #25) for the cadets of ADFA. This course represented the basis for the development of a new Master Program that Dr Guha, Dr Cappuccio and Dr Galliot intended to implement at ADFA. This Master program has never been implemented, but our group members have capitalized on the underlying reflection when they designed and taught the online graduate course of “Cyber security ethics” for the Cybersecurity centre of ADFA. Dr Cappuccio has capitalized on the work done as part of this project to develop an original methodological approach to the use of robots in teaching and education, as testified by his contribution to the book “Robots in Education” (Routledge). Despite the significant difference of scope and domains, the key concerns are the same: technological alienation and the dangers of intolerance for autonomy.

Describe the impact in this reporting period on physical, institutional, and information resources that form infrastructure.

For example: Physical resources such as facilities, laboratories, or instruments; institutional resources (such as establishment or sustenance of societies or organizations); or information resources, electronic means for accessing such resources or for scientific communication, or the like.

Due to its groundbreaking approach, and the fact it has spearheaded an entirely new kind of interdisciplinary research, interpreting an important academic trend and responding to an objective demand from the industry, our project has directly and indirectly influenced similar research efforts conducted in Australia by other groups and institutions with which some of our team members have worked: these include the Trusted Autonomy group at the School of Engineering and Information Technology of ADFA and the Trusted Autonomous Systems Cooperative Research Centre, with which our group has interacted and collaborated for a period. The influence exerted by our project is subtle and largely intangible in nature, in the sense that it is hardly quantifiable in terms of resources or supporting infrastructure, but its depth is attested by the increasing propensity of academic engineers and technology makers working for the industry to consult philosophers of technology and technology ethicists or even hire them among their ranks.

Impact on society beyond science and technology:

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.

In addition to the research papers published in journals and volumes, our work has reached out to the broader society through newspaper articles, online news outlets, social media, and national TV (dr. Galliot was interviewed by ABC). Our group, and specifically the research papers written under the Minerva project, had an impact on the policy decisions made by influential international institutions involved with the regulation of autonomous weapon systems. To begin with, both Dr Natalia Jevglevskaja and Dr Jai Galliot have represented the University of New South Wales at the Australian Defence Force Academy during the 2019 *United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, held on 25-29 March 2019 in Geneva, Switzerland. Dr Jai Galliot has represented UNSW at the 2019 session of the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems in Geneva (20-21 August 2019), held under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. This meeting provided a forum for States and other interested parties to discuss questions related to emerging technologies in the area of lethal autonomous weapons. Dr Natalia Jevglevskaja has presented the paper "Lethal Autonomous Weapons systems", at the Australian Red Cross – Defence Legal Panel on Lethal Autonomous Weapons Systems in Canberra on the 13th of August 2019. Also, Dr. Natalia Jevglevskaja presented the paper "Legal Review of Emerging Military Technology, the Case of Cyber Capabilities and Autonomous Weapons Systems" at the 27th Annual Conference of the Australian & New Zealand Society of International Law, in Canberra, on the 6th of July 2019.

Changes

In this section, please incorporate any and all changes that you would like your program officers to know about the grant. The principal investigator is reminded that the recipient organization is required to obtain prior written approval from program officers whenever there are significant changes in the project or its direction. Sections may include:

Changes in approach

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

The project did not have any significant changes in objectives, scope, or approach.

Problems or delays

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

The project has been afflicted by several problems and unexpected contingencies.

During the four-year project's span, some team members (including, at different times, Dr. Tim McFarland, Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha) resigned from their role in the Minerva-sponsored project either because they had to take on a new job in another institution (Dr. Tim McFarland) or because they had the opportunity to start higher-rank functions at UNSW-ADFA under other externally-funded research projects with different research responsibilities (Dr. Bianca Baggiarini, Dr. Massimiliano Cappuccio, and Dr. Manabrata Guha). Career mobility has created only minor disruptions because the staff turnover has happened at a slow pace and with sufficient notice, but it has created the constant need to re-allocate some of the relevant functions and roles.

The Covid-19 pandemics and the extensive periods of lockdown associated with it have caused significant disruptions in the activities of our group. Especially the organization of public events and the quality and quantity of the communication with the relevant members and associations within the scientific community (for example through social media and newsletters) have been seriously impacted during the period 2020-2021. The collection of empirical data became extremely difficult due to the impossibility to meet with human subjects and obtain approval. Other individual and group activities, such as independent research and collective manuscript preparation have been significantly impacted due to the difficulties associated with working from home, especially for the researchers with children, and the impossibility of holding regular face-to-face meetings. Also, all the researchers involved in our project have experienced a noticeable increase in the average time required by the editors of peer-reviewed journals to organize review and publication of research papers.

Expenditure Impacts

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report.

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Nothing to report.

Changes to the primary place of performance from that originally proposed

Identify any change to the primary performance site location identified in the proposal, as originally submitted.

The project members have been using the rooms and facilities provided by the School of Engineering and Information Technology. As the main space assigned to our team, room 370-71 in Building 21 of the ADFA campus (aka UNSW Canberra) in Campbell ACT has been the primary place of performance throughout the whole project. However, since April 2020, most of the group activities had to be carried out remotely or from home due the disruptions brought by Covid-19.

Technical Updates

This section will include any and all technical updates that you would like to provide to your program officer. You are encouraged to upload graphs and other visualizations that highlight the work done during this reporting period. Program Officers may request additional information that is specific to your research topics.

Nothing to report.