Autonomy and Artificial Intelligence: A Threat or Savior?

W.F. Lawless • Ranjeev Mittu • Donald Sofge
Stephen Russell
Editors

Autonomy and Artificial Intelligence: A Threat or Savior?

Springer

Editors
W.F. Lawless
Paine College
Augusta, GA, USA

Ranjeev Mittu
Naval Research Laboratory
Washington, DC, USA

Donald Sofge
Naval Research Laboratory
Washington, DC, USA

Stephen Russell
U.S. Army Research Laboratory
Adelphi, MD, USA

ISBN 978-3-319-59718-8 ISBN 978-3-319-59719-5 (eBook)
DOI 10.1007/978-3-319-59719-5

Library of Congress Control Number: 2017947297

© Springer International Publishing AG 2017

Chapters 1, 4, 5, 12, and 13 were created within the capacity of an US governmental employment. US copyright protection does not apply.

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book derives from two Association for the Advancement of Artificial Intelligence (AAAI) symposia; the first symposium on "Foundations of Autonomy and Its (Cyber) Threats—From Individuals to Interdependence" was held at Stanford University from March 23 to 25, 2015, and the second symposium on "AI and the Mitigation of Human Error—Anomalies, Team Metrics and Thermodynamics" was held again at Stanford University from March 21 to 23, 2016. This book, titled *Autonomy and Artificial Intelligence: A Threat or Savior?*, combines and extends the themes of both symposia. Our goal for this book is to deal with the current state of the art in autonomy and artificial intelligence by examining the gaps in the existing research that must be addressed to better integrate autonomous and human systems. The research we present in this book will help to advance the next generation of systems that are already planned ranging from autonomous platforms and machines to teams of autonomous systems to provide better support to human operators, decision-makers, and the society.

This book explores how artificial intelligence (AI), by leading to an increase in the autonomy of machines and robots, is offering opportunities for an expanded but uncertain impact on society by humans, machines, and robots. To help readers better understand the relationships between AI, autonomy, humans, and machines that will help society reduce human errors in the use of advanced technologies (e.g., airplanes, trains, cars), this edited volume presents a wide selection of the underlying theories, computational models, experimental methods, and field applications. While other books deal with these topics individually, this book is unique in that it unifies the fields of autonomy and AI and frames them in the broader context of effective integration for human-autonomous machine and robotic systems.

The **introduction** in this volume begins by describing the current state of the art for research in AI, autonomy, and cyber-threats presented at Stanford University in the spring of 2015 (copies of the technical articles are available from AAAI at http://www.aaai.org/Symposia/Spring/sss15symposia.php#ss03; a link to the agenda for the symposium in 2015 along with contact information for the invited speakers and regular participants is at https://sites.google.com/site/foundationsofautonomy-aaais2015/) and for research in AI, autonomy, and error mitigation presented at the

same university in the spring of 2016 (copies of the technical articles are available from AAAI at http://www.aaai.org/Symposia/Spring/sss16symposia.php#ss01; a link to the agenda and contact information for the invited speakers and regular participants is at https://sites.google.com/site/aiandthemitigationofhumanerror/).

After introducing the themes in this book and the contributions from world-class researchers and scientists, individual chapters follow where they elaborate on key research topics at the heart of effective human-machine-robot-systems integration. These topics include computational support for intelligence analyses; the challenge of verifying today's and future autonomous systems; comparisons between today's machines and autism; implications of human-information interaction on artificial intelligence and errors; systems that reason; the autonomy of machines, robots, and buildings; and hybrid teams, where hybrid reflects arbitrary combinations of humans, machines, and robots.

The contributions to this volume are written by leading scientists across the field of autonomous systems research, ranging from industry and academia to government. Given the broad diversity of the research in this book, we strove to thoroughly examine the challenges and trends of systems that implement and exhibit AI; social implications of present and future systems made autonomous with AI; systems with AI seeking to develop trusted relationships among humans, machines, and robots; and effective human systems integration that must result for trust in these new systems and their applications to increase and to be sustained.

A brief summary of the AAAI symposia in the spring of 2015 and the spring of 2016 is presented below.

Spring 2015: Foundations of Autonomy and Its (Cyber) Threats—From Individuals to Interdependence

Spring 2015: Organizing Committee

Ranjeev Mittu (ranjeev.mittu@nrl.navy.mil), Naval Research Laboratory
Gavin Taylor (taylor@usna.edu), US Naval Academy
Donald Sofge (don.sofge@nrl.navy.mil), Naval Research Laboratory, Navy Center for Applied Research in Artificial Intelligence
William F. Lawless (wlawless@paine.edu), Paine College, Departments of Math and Psychology

Spring 2015: Program Committee

- David Atkinson (datkinson@ihmc.us), Senior Research Scientist, Florida Institute for Human and Machine Cognition

- Lashon B. Booker (booker@mitre.org), Ph.D., Senior Principal Scientist, The MITRE Corporation
- Jeffery Bradshaw (jbradshaw@ihmc.us), Senior Research Scientist, Florida Institute for Human and Machine Cognition
- Michael Floyd (michael.floyd@knexusresearch.com), Knexus Research
- Sharon Graves (sharon.s.graves@nasa.gov), NASA Deputy Project Manager, Safe Autonomous Systems Operations, Aeronautics Research Directorate
- Vladimir Gontar (galita@bgu.ac.il), Department of Industrial Engineering and Management, Ben-Gurion University of the Negev
- L. Magafas (lmagafas@otenet.gr), Director of Electronics and Signal Processing Lab., Eastern Macedonia and Thrace Institute of Technology, Kavala, GR
- Bolivar Rocha (bolivar.rocha@gmail.com), Brazil
- Satyandra K. Gupta (skgupta@umd.edu), Director, University of Maryland Robotics Center, Department of Mechanical Engineering and Institute for Systems Research
- Laurent Chaudron (laurent.chaudron@polytechnique.org), Director, ONERA Provence Research Center, French Air Force Academy
- Charles Howell (howell@mitre.org), Chief Engineer for Intelligence Programs and Integration, National Security Engineering Center, The MITRE Corporation
- Jennifer Burke (jennifer.l.burke2@boeing.com), Manager, Human-System Integrated Technologies, Boeing Research and Technology
- Tsuyoshi Murata (murata@cs.titech.ac.jp), Dept. of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology
- Julie Marble (julie.marble@navy.mil), Office of Naval Research, Program Officer for Hybrid Human-Computer Systems
- Doug Riecken (dougriecken@gmail.com), Columbia University Center for Computational Learning Systems
- Catherine Tessier (Catherine.Tessier@onera.fr), Senior Researcher, Dept. of Systems Control and Flight Dynamics, French Aerospace Lab, ONERA, Toulouse, France
- Simon Parsons (s.d.parsons@liverpool.ac.uk), Liverpool, Visiting Professor, Dept. of Computer Science, University of Liverpool; Dept. Graduate Deputy Chair and Co-Dir., Agents Lab, Brooklyn College
- Ciara Sibley (ciara.sibley@nrl.navy.mil), Engineering Research Psychologist, Naval Research Laboratory, Washington, DC

Spring 2015: Invited Keynote Speakers

- Gautam Trivedi (gautam.trivedi@nrl.navy.mil) and Brandon Enochs (brandon.enochs@nrl.navy.mil), Naval Research Laboratory, "Detecting, Analyzing and Locating Unauthorized Wireless Intrusions into Networks"
- Chris Berka (chris@b-alert.com), Advanced Brain Monitoring, "On the Road to Autonomy: Evaluating and Optimizing Hybrid Team Dynamics"

- Kristin E. Schaefer (kristin.e.schaefer2.ctr@mail.mil), US Army Research Lab (ARL), "Perspectives of Trust: Research at the US Army Research Laboratory"
- David R. Martinez (DMartinez@LL.mit.edu), Lincoln Laboratory, Massachusetts Institute of Technology, "Cyber Anomaly Detection with Machine Learning"
- Vladimir Gontar (vgontar@ucsd.edu), BioCircuits Institute, University of California San Diego (UCSD), Ben-Gurion University of the Negev, "Artificial Brain Systems Based on Neural Networks Discrete Chaotic Biochemical Reactions Dynamics and Its Application to Conscious and Creative Robots"

Spring 2015: Regular Speakers

- Christopher A. Miller (cmiller@sift.net), Smart Information Flow Technologies, "Delegation, Intent, Cooperation and Their Failures"
- Ciara Sibley[1] (ciara.sibley@nrl.navy.mil), Joseph Coyne[1] (joseph.coyne@nrl.navy.mil), and Jeffery Morrison[2] (jeffrey.morrison@nrl.navy.mil), [1]Naval Research Laboratory, [2]Office of Naval Research, "Research Considerations for Managing Future Unmanned Systems"
- Gavin Taylor (taylor@usna.edu), Kawika Barabin, and Kent Sayre, Computer Science Department, US Naval Academy, Annapolis, MD 21402-5002, "An Application of Reinforcement Learning to Supervised Autonomy"
- David J. Atkinson (datkinson@ihmc.us), Florida Institute for Human and Machine Cognition, Ocala, FL, "Emerging Cyber-Security Issues of Autonomy and the Psychopathology of Intelligent Machines"
- Olivier Bartheye[1] (olivier.barteye@intradef.gouv.fr) and Laurent Chaudron[2] (laurent.chaudron@polytechnique.org), CREC St-Cyr[1] and ONERA[2], "Risk Management Systems Must Provide Automatic Decisions for Crisis Computable Algebras"
- William F. Lawless (wlawless@paine.edu), Paine College, Augusta, GA, and Ira S. Moskowitz, Ranjeev Mittu, and Donald A. Sofge ({ira.moskowitz; ranjeev.mittu; donald.sofge}@nrl.navy.mil), Naval Research Laboratory, Washington, DC, "A Thermodynamics of Teams: Towards a Robust Computational Model of Autonomous Teams"
- Ranjeev Mittu[1] (ranjeev.mittu@nrl.navy.mil) and Julie Marble[2] (julie.marble@navy.mil), [1]Naval Research Laboratory, Information Technology Division, Washington, DC; [2] Office of Naval Research, VA 22203-1995 (changing to Johns Hopkins Applied Physics Lab, MD), "The Human Factor in Cybersecurity: Robust and Intelligent Defense"
- Myriam Abramson (myriam.abramson@nrl.navy.mil), Naval Research Laboratory, Washington, DC, "Cognitive Fingerprints"
- Ira S. Moskowitz[1] (ira.moskowitz@nrl.navy.mil), William F. Lawless[2], (wlawless@paine.edu), Paul Hyden[1] (paul.hyden@nrl.navy.mil), Ranjeev Mittu[1] (ran-

jeev.mittu@nrl.navy.mil), and Stephen Russell[1] (stephen.m.russell8.civ@mail.mil), [1]Information Management and Decision Architectures Branch, Naval Research Laboratory, Washington, DC; [2]Departments of Mathematics and Psychology, Paine College, Augusta, GA, "A Network Science Approach to Entropy and Training"
- Boris Galitsky (bgalitsky@hotmail.com), Knowledge Trail Inc., San Jose, CA, "Team Formation by Children with Autism"
- Olivier Bartheye[1] (olivier.barteye@intradef.gouv.fr) and Laurent Chaudron[2] (laurent.chaudron@polytechnique.org), CREC St-Cyr[1] and ONERA[2], "Algebraic Models of the Self-Orientation Concept for Autonomous Systems"

Spring 2016: AI and the Mitigation of Human Error—Anomalies, Team Metrics and Thermodynamics

Spring 2016: Organizing Committee

Ranjeev Mittu (ranjeev.mittu@nrl.navy.mil), Naval Research Laboratory
Gavin Taylor (taylor@usna.edu), US Naval Academy
Donald Sofge (don.sofge@nrl.navy.mil), Naval Research Laboratory
William F. Lawless (wlawless@paine.edu), Paine College, Departments of Math and Psychology

Spring 2016: Program Committee (duplicates the spring 2015 symposium)

Spring 2016: Invited Keynote Speakers

- Julie Adams (julie.a.adams@vanderbilt.edu), Vanderbilt University, Associate Professor of Computer Science and Computer Engineering, Electrical Engineering and Computer Science Department, "AI and the Mitigation of Error"
- Stephen Russell (stephen.m.russell8.civ@mail.mil), Chief, Battlefield Information Processing Branch, US Army Research Lab, MD, "Human Information Interaction, Artificial Intelligence, and Errors"
- James Llinas (llinas@buffalo.edu), SUNY at Buffalo, "An Argumentation-Based System Support Toolkit for Intelligence Analyses"
- Martin Voshell (mvoshell@cra.com), Charles River Analytics, "Multi-Level Human-Autonomy Teams for Distributed Mission Management"

Spring 2016: Regular Speakers

- Ira S. Moskowitz (ira.moskowitz@nrl.navy.mil), NRL; "Human-Caused Bifurcations in a Hybrid Team—A Position Paper"
- Paul Hyden (paul.hyden@nrl.navy.mil), NRL, "Fortification Through Topological Dominance: Using Hop Distance and Randomized Topology Strategies to Enhance Network Security"
- Olivier Bartheye (olivier.barteye@intradef.gouv.fr), CREC St-Cyr, , and Laurent Chaudron (laurent.chaudron@polytechnique.org), ONERA, "Epistemological Qualification of Valid Action Plans for UGVs or UAVs in Urban Areas"
- William F. Lawless, (wlawless@paine.edu), Paine College, "AI and the Mitigation of Error: A Thermodynamics of Teams"

Questions for Speakers and Attendees at AAAI-2015 and AAAI-2016 and for Readers of This Book

Our spring AAAI-2015 and AAAI-2016 symposia offered speakers opportunities with AI to address the intractable, fundamental questions about cybersecurity, machines and robots, autonomy and its management, the malleability of preferences and beliefs in social settings, or the application of autonomy for hybrids at the individual, group, and system levels.

A list of unanswered fundamental questions included:

- Why have we yet to determine from a theoretical perspective the principles underlying individual, team, and system behaviors?
- Can autonomous systems be controlled to solve the problems faced by teams while maintaining defenses against threats and minimizing mistakes in competitive environments (e.g., cyber attacks, human error, system failure)?
- Do individuals seek to self-organize into autonomous groups like teams in order to better defend against attacks (e.g., cyber, merger, resources) or for other reasons (e.g., least entropy production (LEP) and maximum entropy production (MEP))?
- What does an autonomous organization need to predict its path forward and govern itself? What are the AI tools available to help an organization be more adept and creative?
- What signifies adaptation? For AI, does adaptation at an earlier time prevent or moderate adaptive responses to newer environmental changes?
- Is the stability state of hybrid teams the single state that generates the MEP rate?
- If social order requires MEP, and if the bistable perspectives present in debate (courtrooms, politics, science) lead to stable decisions, is the chosen decision an LEP or MEP state?

- Considering the evolution of social systems (e.g., in general, Cuba, North Korea, and Palestine have not evolved), are the systems that adjust to MEP the most efficient?

In addition, new threats may emerge due to the nature of the technology of autonomy itself (as well as the breakdown in traditional verification and validation (V&V) and test and evaluation (T&E) due to the expanded development and application of AI). This nature of advanced technology leads to other key AI questions for consideration now and in the future:

Fault Modes

- Are there new types of fault modes that can be exploited by outsiders?

Detection

- How can we detect that an intelligent, autonomous system has been or is being subverted?

Isolation

- What is a "fail-safe" or "fail-operational" mode for an autonomous system, and can it be implemented?
- Implication of cascading faults (AI, system, cyber)

Resilience and Repair

- What are the underlying causes of the symptoms of faults (e.g., nature of the algorithms, patterns of data, etc.)?

Consequences of Cyber Vulnerabilities

- Inducement of fault modes
- Deception (including false flags)
- Subversion
- The human/social element (reliance, trust, and performance)

We invited speakers and attendants at our two symposia to address the following more specific AI topics (as we invite readers of this book to consider):

- Computational models of autonomy (with real or virtual individuals, teams, or systems) and performance (e.g., metrics, MEP) with or without interdependence, uncertainty, and stability
- Computational models that address autonomy and trust (e.g., the trust by autonomous machines of human behavior or the trust by humans of autonomous machine behavior)
- Computational models that address threats to autonomy and trust (cyber attacks, competitive threats, deception) and the fundamental barriers to system survivability (e.g., decisions, mistakes, etc.)

- Computational models for the effective or efficient management of complex systems (e.g., the results of decision-making, operational performance, metrics of effectiveness, efficiency)
- Models of multi-agent systems (e.g., multi-UAVs, multi-UxVs, model verification and validation) that address autonomy (e.g., its performance, effectiveness, and efficiency).

For future research projects and symposia (e.g., our symposium in 2017 on "Computational Context: Why It's Important, What It Means, and Can It Be Computed?"; see http://www.aaai.org/Symposia/Spring/sss17symposia.php#ss03), we invite readers to consider other questions or topics from individual (e.g., cognitive science, economics), machine learning (ANNs; GAs), or interdependent (e.g., team, firm, system) perspectives.

After the AAAI-spring symposia in 2015 and 2016 were completed, the symposia presentations and technical reports and the book took on separate lives. The following individuals were responsible for the proposal submitted to Springer after the symposia, for the divergence between the topics considered by the two, and for editing this book that has resulted:

Augusta, GA, USA W.F. Lawless
Washington, DC, USA Ranjeev Mittu
Adelphi, MD, USA Donald Sofge
Washington, DC, USA Stephen Russell

Contents

1 Introduction 1
W.F. Lawless, Ranjeev Mittu, Stephen Russell, and Donald Sofge

2 Reexamining Computational Support for Intelligence Analysis: A Functional Design for a Future Capability 13
James Llinas, Galina Rogova, Kevin Barry, Rachel Hingst, Peter Gerken, and Alicia Ruvinsky

3 Task Allocation Using Parallelized Clustering and Auctioning Algorithms for Heterogeneous Robotic Swarms Operating on a Cloud Network 47
Jonathan Lwowski, Patrick Benavidez, John J. Prevost, and Mo Jamshidi

4 Human Information Interaction, Artificial Intelligence, and Errors 71
Stephen Russell, Ira S. Moskowitz, and Adrienne Raglin

5 Verification Challenges for Autonomous Systems 103
Signe A. Redfield and Mae L. Seto

6 Conceptualizing Overtrust in Robots: Why Do People Trust a Robot That Previously Failed? 129
Paul Robinette, Ayanna Howard, and Alan R. Wagner

7 Research Considerations and Tools for Evaluating Human-Automation Interaction with Future Unmanned Systems 157
Ciara Sibley, Joseph Coyne, and Sarah Sherwood

8 Robots Autonomy: Some Technical Issues 179
Catherine Tessier

9 How Children with Autism and Machines Learn to Interact 195
Boris A. Galitsky and Anna Parnis

10 Semantic Vector Spaces for Broadening Consideration of Consequences 227
Douglas Summers-Stay

11 On the Road to Autonomy: Evaluating and Optimizing Hybrid Team Dynamics 245
Chris Berka and Maja Stikic

12 Cybersecurity and Optimization in Smart "Autonomous" Buildings 263
Michael Mylrea and Sri Nikhil Gupta Gourisetti

13 Evaluations: Autonomy and Artificial Intelligence: A Threat or Savior? 295
W.F. Lawless and Donald A. Sofge