Artificial Intelligence Safety and Security

Call for Book Chapters

Introduction

The history of robotics and artificial intelligence in many ways is also the history of humanity's attempts to control such technologies. From the Golem of Prague to the military robots of modernity, the debate continues as to what degree of independence such entities should have and how to make sure that they do not turn on us, its inventors. Numerous recent advancements in all aspects of research, development and deployment of intelligent systems are well publicized but safety and security issues related to AI are rarely addressed. This book is proposed to mitigate this fundamental problem. It will be comprised of chapters from leading AI Safety researchers addressing different aspects of the AI control problem as it relates to the development of safe and secure artificial intelligence. The book would be the first textbook to address challenges of constructing safe and secure advanced machine intelligence.

Contributions devoted to **all** aspects of AI safety and security are invited, in particular:

- Adversarial ML
- AI Failures
- Bad Actors/Malevolence
- Boxing/Confinement
- Drives/Goals
- Ethics/Law

- Explainability
- Friendliness Theory
- Openness/Secrecy
- Off-Switch
- Reward Engineering
- Self-Improvement

- Security/Control
- Singularity/Superintelligence
- Validation/Verification
- Value Alignment
- Weaponized/Military
- Wireheading

Submission Procedure

Researchers and practitioners are invited to submit their papers by email. Chapters and inquiries should be submitted to Dr. Roman Yampolskiy at roman.yampolskiy@louisville.edu. Submissions should include authors' information (names, affiliations, emails, and short bios). Note: There are NO submission or acceptance fees for manuscripts submitted to this book for publication. All manuscripts are accepted based on a double-blind peer review editorial process. Contributors may also be requested to serve as reviewers for this project. Authors of accepted chapters will get a choice of free e-book or printed book after publication.

Important Dates:

Chapter Submission: before July 25, 2017;
Reviews and Decisions are due: September 20, 2017;
Revised Chapters are due: October 31, 2017. This book is anticipated to be released in early 2018.

Formatting Instructions

A chapter is recommended to have about 25 double-spaced A4 pages with 1" margins (longer or shorter chapters are completely fine). Chapters should be proofread, polished and formatted according to the publisher's instructions before submission. We will deliver camera-ready chapters of the book to the publisher. Author's guidelines and chapter templates can be found here (https://goo.gl/d3LJuj).

Editor: Dr. Roman V. Yampolskiy, University of Louisville, USA. roman.yampolskiy@louisville.edu.
    Author of *Artificial Superintelligence* (CRC Press) and co-Editor of The Technological Singularity II (Springer).