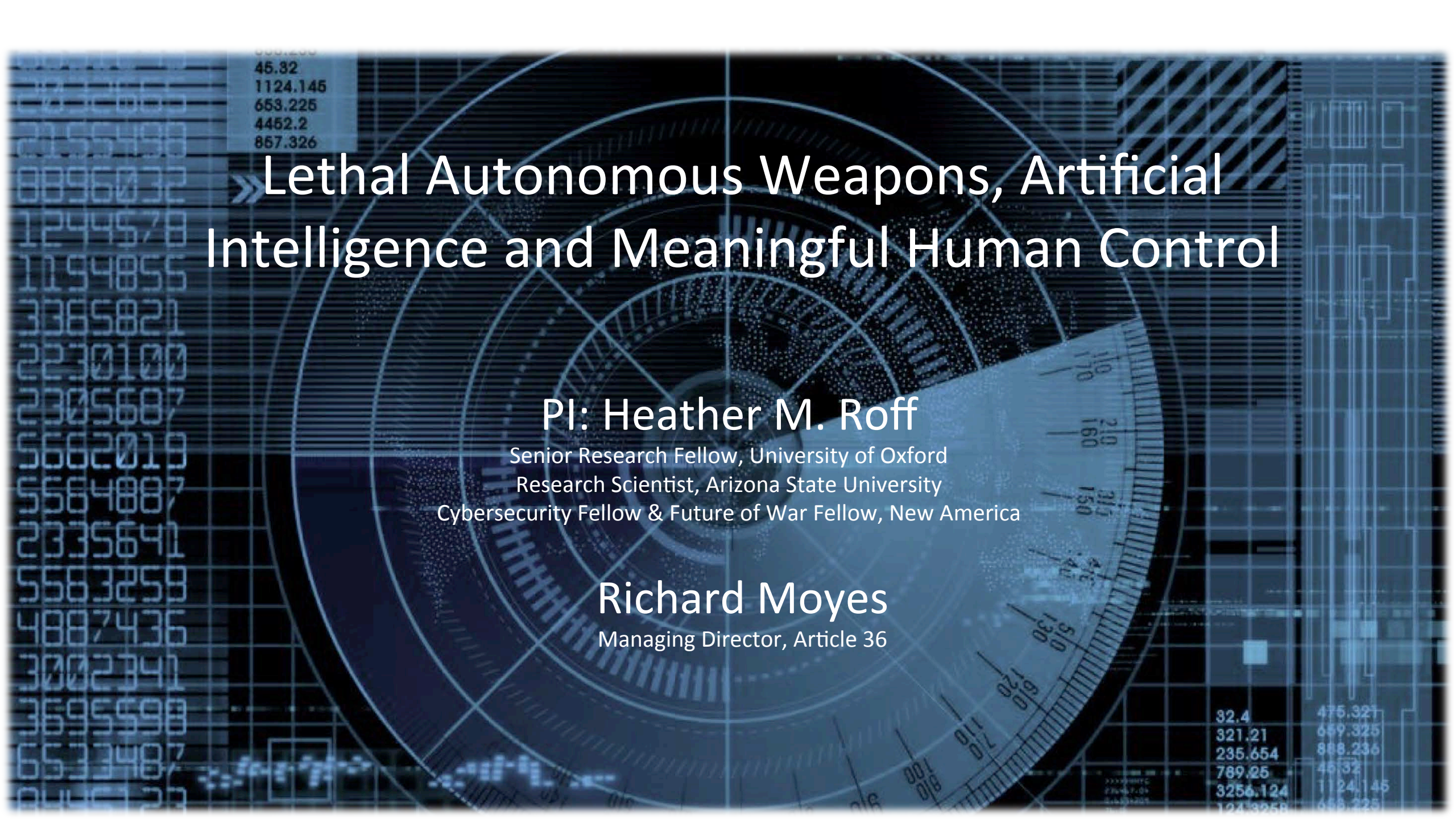


Lethal Autonomous Weapons, Artificial Intelligence and Meaningful Human Control

PI: Heather M. Roff

Senior Research Fellow, University of Oxford
Research Scientist, Arizona State University
Cybersecurity Fellow & Future of War Fellow, New America

Richard Moyes

Managing Director, Article 36

The Project: 1 Data

Create a dataset on presently deployed weapons systems in top 5 weapons exporting countries. Code all automated capacities on such systems. To date: N = 256

Cases:

- USA (31%)
- Russia (27%)
- China* (5%)
- Germany (5%)
- France (5%)

Largest Weapons Manufacturers within Each Country:

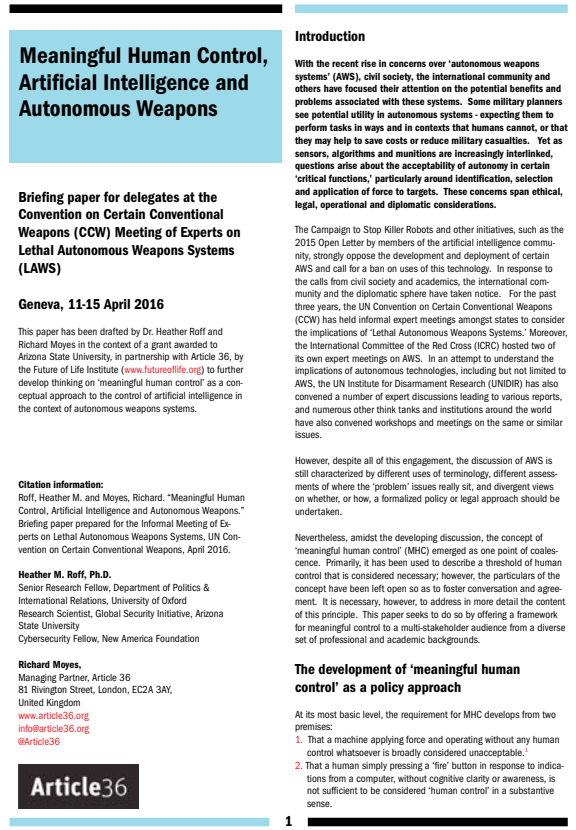
- Data from Stockholm International Peace Research Institute, Top 100 Arms Producing Companies
- Publically available information on all systems produced by each of these companies

The Project: 2 Concepts

Create a conceptual framework for the concept of “Meaningful Human Control;” define values associated with it.

Multi-stakeholder approach:

- Academia: robotics, AI, philosophy, cognitive science, psychology
- Military Lawyers; Military Officers, Joint Chiefs (US)
- Governments (official representations for disarmament issues)
- NGOs (civil society)
- United Nations Institute for Disarmament Research (UNIDIR)
- International Committee for the Red Cross (ICRC)



Meaningful Human Control, Artificial Intelligence and Autonomous Weapons

Briefing paper for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)

Geneva, 11-15 April 2016

This paper has been drafted by Dr. Heather Roff and Richard Moyes in the context of a grant awarded to Arizona State University, in partnership with Article 36, by the Future of Life Institute (www.futureoflife.org) to further develop thinking on ‘meaningful human control’ as a conceptual approach to the control of artificial intelligence in the context of autonomous weapons systems.

Citation information:
Roff, Heather M. and Moyes, Richard. “Meaningful Human Control, Artificial Intelligence and Autonomous Weapons.” Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons, April 2016.

Heather M. Roff, Ph.D.
Senior Research Fellow, Department of Politics & International Relations, University of Oxford
Research Scientist, Global Security Initiative, Arizona State University
Cybersecurity Fellow, New America Foundation

Richard Moyes,
Managing Partner, Article 36
81 Rivington Street, London, EC2A 3AY,
United Kingdom
www.article36.org
info@article36.org
@Article36

Article36

Introduction

With the recent rise in concerns over ‘autonomous weapons systems’ (AWS), civil society, the international community and others have focused their attention on the potential benefits and problems associated with these systems. Some military planners see potential utility in autonomous systems - expecting them to perform tasks in ways and in contexts that humans cannot, or that they may help to save costs or reduce military casualties. Yet as sensors, algorithms and munitions are increasingly interlinked, questions arise about the acceptability of autonomy in certain ‘critical functions,’ particularly around identification, selection and application of force to targets. These concerns span ethical, legal, operational and diplomatic considerations.

The Campaign to Stop Killer Robots and other initiatives, such as the 2015 Open Letter by members of the artificial intelligence community, strongly oppose the development and deployment of certain AWS and call for a ban on uses of this technology. In response to the calls from civil society and academics, the international community and the diplomatic sphere have taken notice. For the past three years, the UN Convention on Certain Conventional Weapons (CCW) has held informal expert meetings amongst states to consider the implications of ‘Lethal Autonomous Weapons Systems.’ Moreover, the International Committee of the Red Cross (ICRC) hosted two of its own expert meetings on AWS. In an attempt to understand the implications of autonomous technologies, including but not limited to AWS, the UN Institute for Disarmament Research (UNIDIR) has also convened a number of expert discussions leading to various reports, and numerous other think tanks and institutions around the world have also convened workshops and meetings on the same or similar issues.

However, despite all of this engagement, the discussion of AWS is still characterized by different uses of terminology, different assessments of where the ‘problem’ issues really sit, and divergent views on whether, or how, a formalized policy or legal approach should be undertaken.

Nevertheless, amidst the developing discussion, the concept of ‘meaningful human control’ (MHC) emerged as one point of coalescence. Primarily, it has been used to describe a threshold of human control that is considered necessary; however, the particulars of the concept have been left open so as to foster conversation and agreement. It is necessary, however, to address in more detail the content of this principle. This paper seeks to do so by offering a framework for meaningful control to a multi-stakeholder audience from a diverse set of professional and academic backgrounds.

The development of ‘meaningful human control’ as a policy approach

At its most basic level, the requirement for MHC develops from two premises:

1. That a machine applying force and operating without any human control whatsoever is broadly considered unacceptable.’
2. That a human simply pressing a ‘fire’ button in response to indications from a computer, without cognitive clarity or awareness, is not sufficient to be considered ‘human control’ in a substantive sense.

The Project: 3 Impact

International Policy Development
International Norm Generation

Défense nationale National Defence

The Outputs (besides the data)

- **Policy Impact:**

- Dr. Roff & Mr. Moyes Testify separately to ICRC Meeting of Experts on Lethal Autonomous Weapons (March)
- Dr. Roff & Mr. Moyes Testify separately to Informal Meeting of Experts on Lethal Autonomous Weapons at the UN Convention on Conventional Weapons (April)
- Dr. Roff presents at the UK MoD's International Weapons Review Forum (October)
- Dr. Roff & Mr. Moyes speak at side event at UN CCW Review Conference hosted by Canada (December)

- **Policy Papers:**
  - Heather M. Roff & Richard Moyes. "Meaningful Human Control, Lethal Autonomous Weapons and Artificial Intelligence" Briefing for UN CCW delegates
  - Richard Moyes. "Article 36 Reviews and Addressing Lethal Autonomous Weapon Systems"
  - Richard Moyes. "The United Kingdom and Autonomous Weapons Systems"

- **Academic Papers:**
  - "An Ontology of Autonomy: Autonomy in Weapons Systems" in *The Ethics of Autonomous Weapons*, edited by Claire Finkelstein, Duncan MacIntosh, and Jens David Ohlin (Oxford University Press, forthcoming).
  - **"**Advancing Human Security Through Artificial Intelligence" in *Emerging Technologies and Human Security*, (Chatham House, forthcoming)
  - **"**The Necessity and Limits of Trust in Autonomous Weapons Systems" Co-authored with David Danks, (under review)
  - "The Forest for the Trees: Autonomous Weapons or Autonomy in Weapons Systems" (work in progress)

- **Popular Press:**
  - Heather M. Roff. "Civilian Harm and Lethal Autonomous Weapons" *The Bulletin of the Atomic Scientists* (November 2015-2016)
  - Heather M. Roff & Peter W. Singer. "The New President Needs a Policy on Autonomous Weapons" *Wired Magazine*, September 6, 2016.
  - Heather M. Roff. "Killer Robots on the Battlefield: the danger of using a war of attrition strategy with autonomous weapons" *Slate Magazine*, April 7, 2016.

The Data: Coverage

Defense One

Report: Weapons AI Increasingly Replacing, Not Augmenting, Human Decision Making

AA FONT SIZE + PRINT AP / STEVE HELBER

BY PATRICK TUCKER
READ BIO

SEPTEMBER 26, 2016

TOPICS

DRONES

A new survey of existing and planned smart weapons finds that AI is increasingly used to replace humans, not help them.

The Pentagon's oft-repeated line on artificial intelligence is this: we need much more of it, and quickly, in order to help humans and machines work better alongside one another. But a survey of existing weapons finds that the U.S. military more commonly uses AI not to help but to replace human operators, and, increasingly, human decision making.

The report from the Elon Musk-funded Future of Life Institute

FP

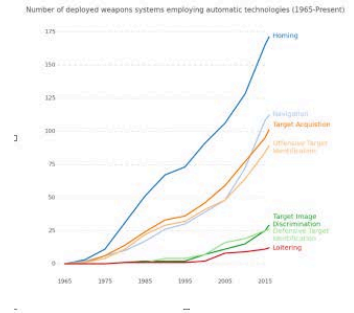
NEWS & IDEAS | REGIONS | CHANNELS | GALLERIES | VOICES

USSR 25th Anniversary | War in Syria | Hacking and Cyber | The Islamic State | ABOUT FP

VOICE

Weapons autonomy is rocketing

BY THOMAS E. RICKS | SEPTEMBER 28, 2016



By Heather M. Roff, Ph.D.
Best Defense guest columnist

While we debate whether or not it is a good idea, weapons are steadily becoming more autonomous, most notably in target identification.

SHARE + 76 SHARES

MILITARY VOICE

About Thomas E. Ricks

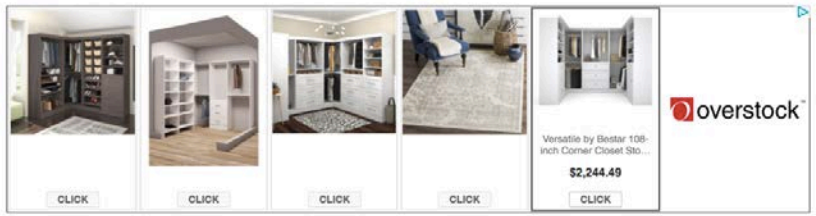
Thomas E. Ricks covered the U.S. military for the Washington Post from 2000 through 2008. He can be reached at ricksblogcomment@gmail.com.

@TOMRICKS1

TRENDING | LATEST | MY FP

All Aboard China's 'New Silk Road' Express
ROBBIE GRAMER | 696 SHARES

POPULAR SCIENCE SUBSCRIBE



MILITARY

AUTONOMOUS WEAPONS ARE ALREADY HERE

A REPORT SHOWS THE EXTENT TO WHICH MACHINES ALREADY MAKE DEADLY DECISIONS

By Kelsey D. Atherton September 28, 2016

SWP

Stiftung Wissenschaft und Politik
German Institute for International and Security Affairs

Meaningful Human Control

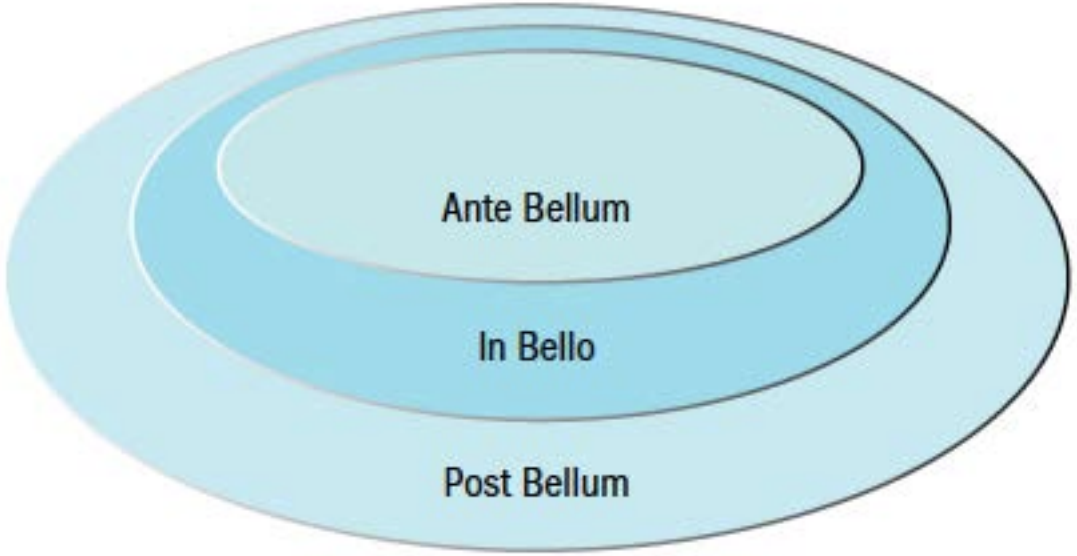


Fig. 1. Human control needs to be embedded through mechanisms operating before, during and after use of technologies in conflict.

- Predictable, reliable and transparent technology.
- Accurate information for the user on the outcome sought, operation and function of technology, and the context of use.
- Timely human action and a potential for timely intervention.
- Accountability to a certain standard

Human control during attacks

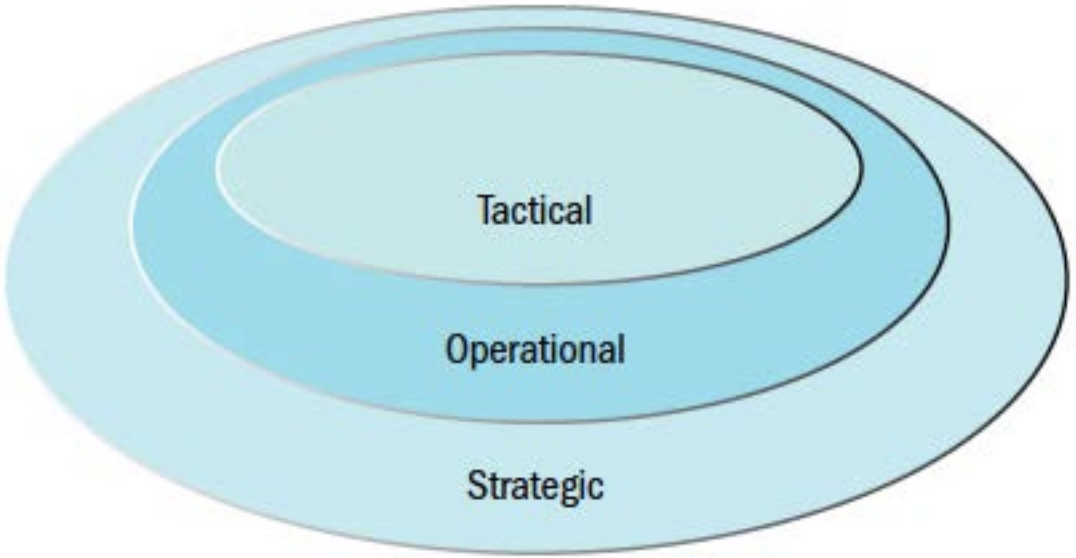
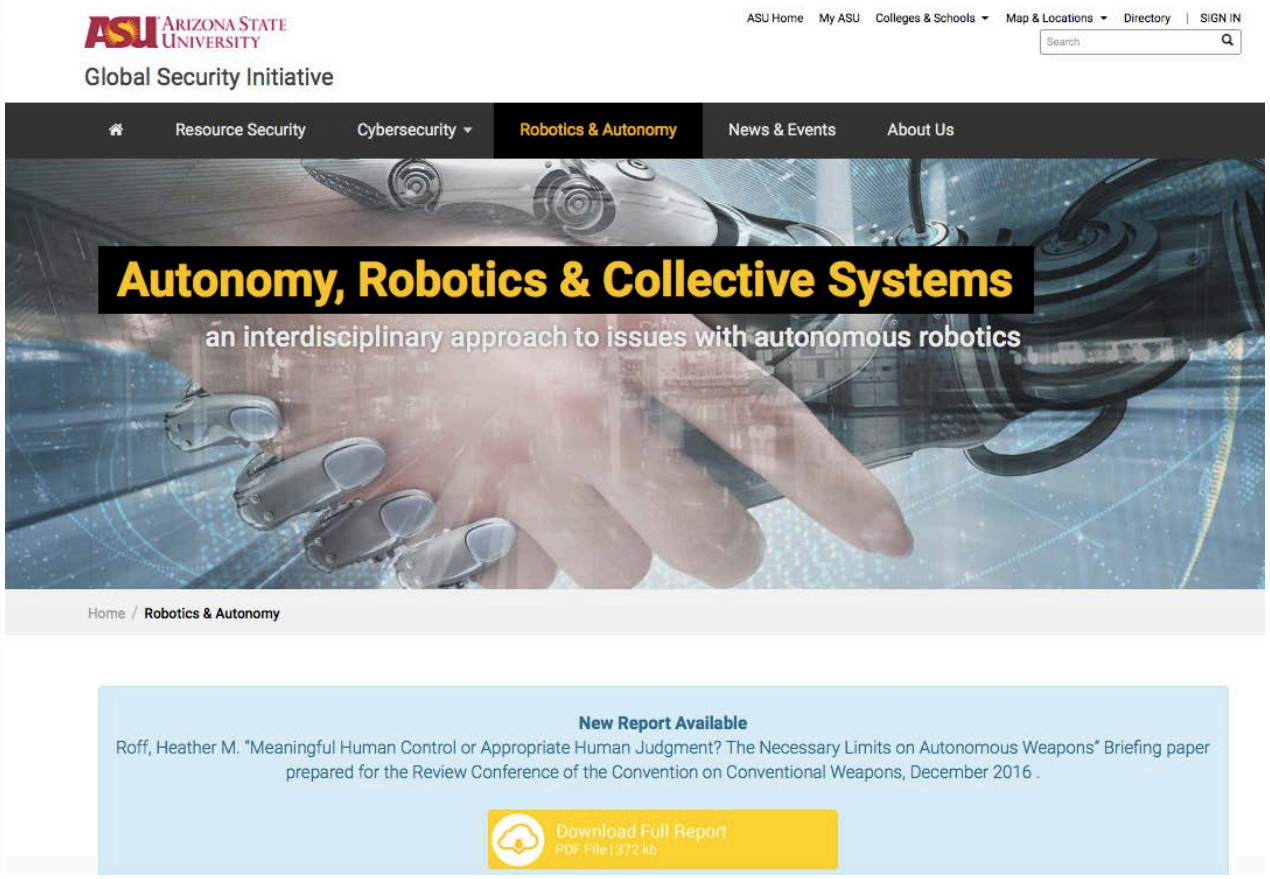


Fig. 2. Meaningful human control needs to be applied over attacks at the tactical level of warfighting, as well at other levels.

The Dataset

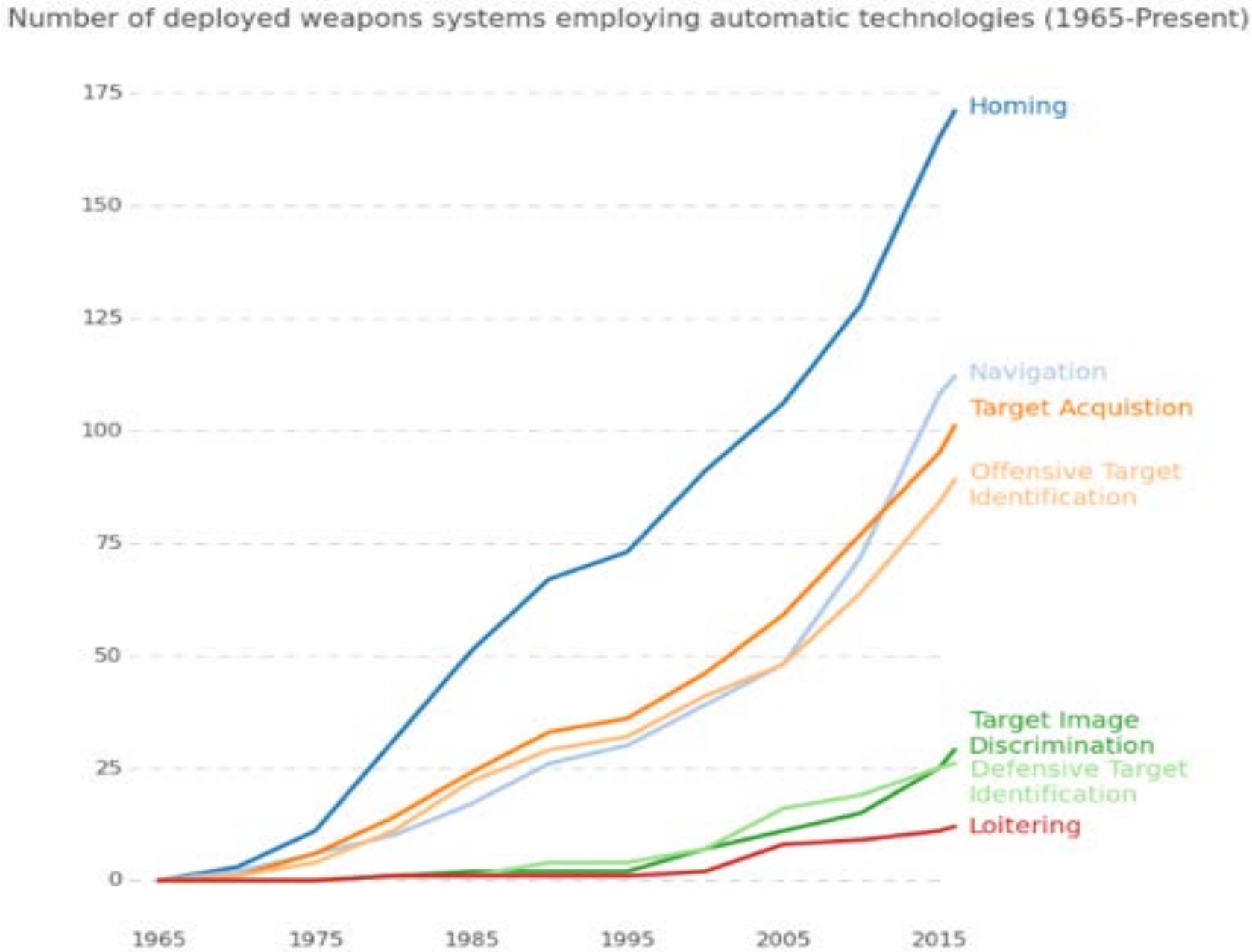


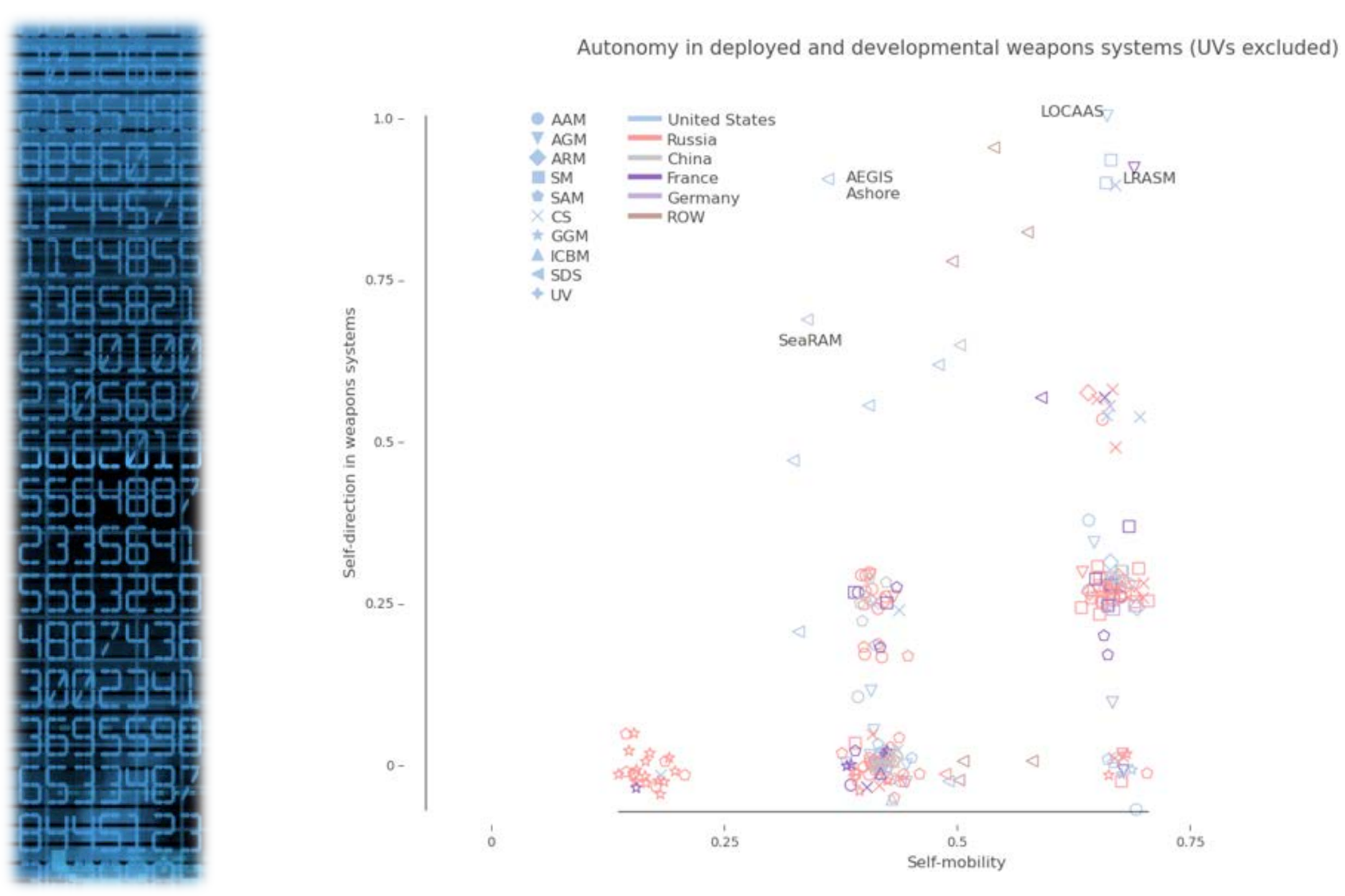
Key Variables

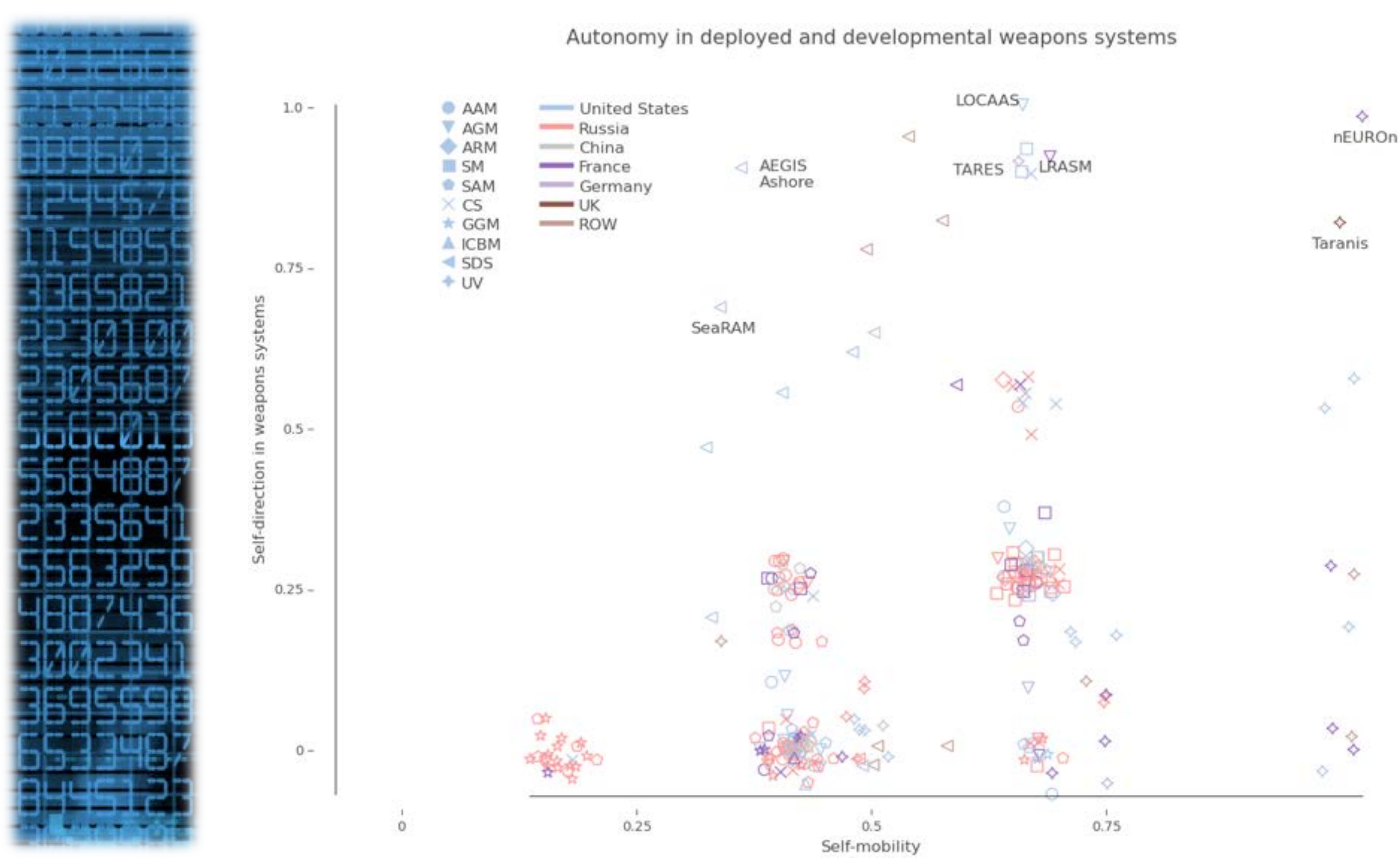
- Mobility (0/1)
- Homing (0/1)
- Navigation (0/1)
- Persistence (0/1)
- Target ID Defensive (0/1)
- Target ID Offensive (0/1)
- Target Image Discrimination (0/1)
- Target Ranking/Priority (0/1)
- Acquire (0/1)
- Fire Control (0-4)
- Engagement Decision (0/1)
- Auto Communication Sharing (0/1)
- Goal Setting/ Planning (0/1)
- Goal Self-Modification (0/1)
- Learning/Adaptation (0/1)

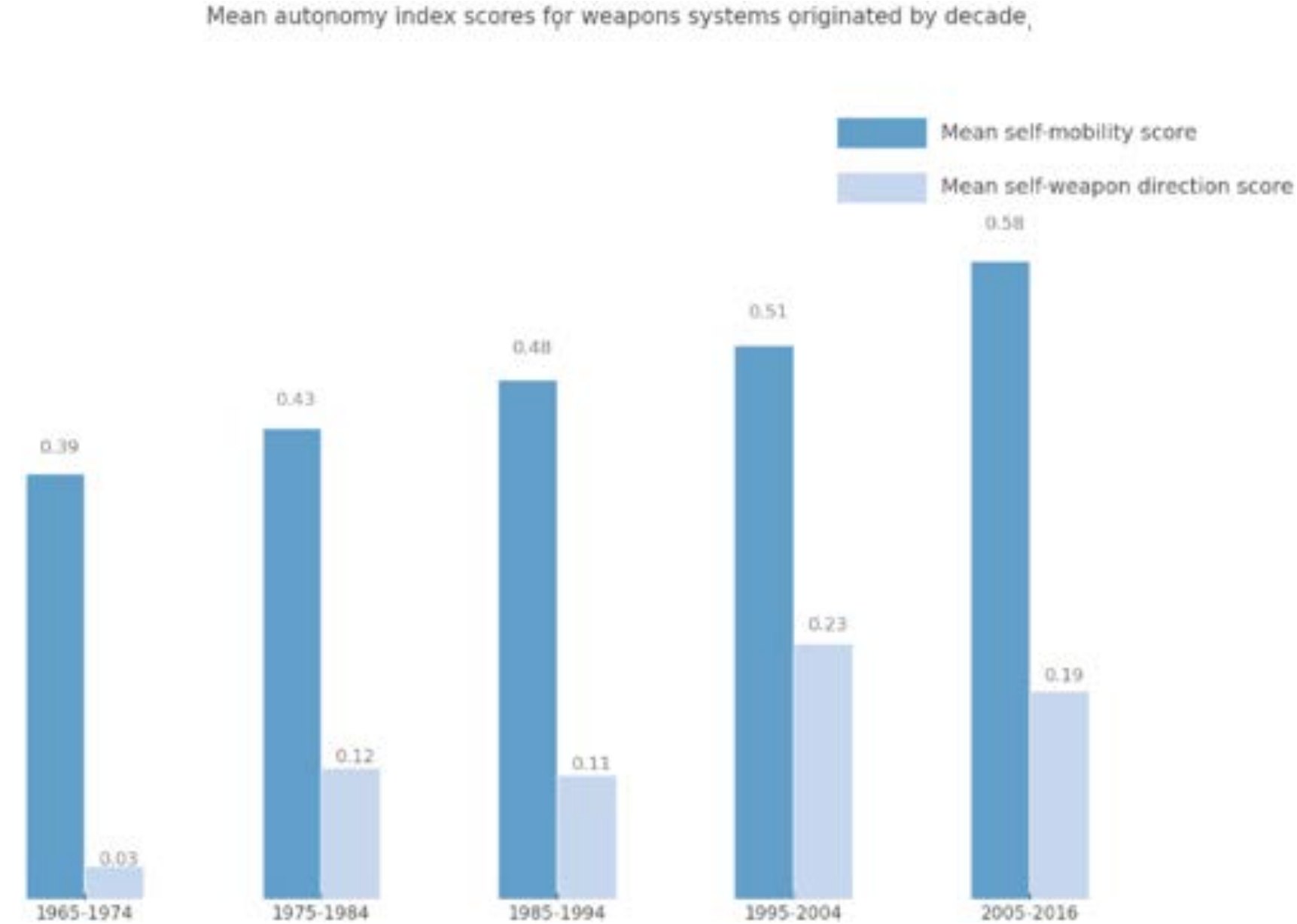
Autonomy Indices

- Condense key variables into three indices
- Each is a weighted, normalized sum of a subset of the key variables
  - 1: Self-Mobility (Mobility, Persistence, Navigation, Homing)
  - 2: Self-Direction in Weapons Systems (Target ID, Target Image Discrimination, Target Ranking/Prioritization, Acquire, Weapon Multi)
  - 3: Self-Determination (Engagement Decision, Auto Communication Sharing, Goal Setting/Planning, Goal Self-Modification, Learning/ Adaptation)
- Scores meant to show trends and relative differences in various capabilities, with 1.0 the max score for each index


Number of deployed weapons systems employing automatic technologies (1965-Present)
175
150
125
100
75
50
25
0
1965
1975
1985
1995
2005
2015
Homing
Navigation
Target Acquistion
Offensive Target Identification
Target Image Discrimination
Defensive Target Identification
Loitering


Autonomy in deployed and developmental weapons systems (UVs excluded)
AAM
AGM
ARM
SM
SAM
CS
GGM
ICBM
SDS
UV
United States
Russia
China
France
Germany
ROW
LOCAAS
LRASM
AEGIS
Ashore
SeaRAM
1.0
0.75
0.5
0.25
0
Self-direction in weapons systems
0
0.25
0.5
0.75
Self-mobility


Autonomy in deployed and developmental weapons systems
AAM
AGM
ARM
SM
SAM
CS
GGM
ICBM
SDS
UV
United States
Russia
China
France
Germany
UK
ROW
LOCAAS
nEUROn
AEGIS
Ashore
TARES
LRASM
Taranis
SeaRAM
1.0
0.75
0.5
0.25
0
Self-direction in weapons systems
0
0.25
0.5
0.75
Self-mobility


Mean autonomy index scores for weapons systems originated by decade
Mean self-mobility score
Mean self-weapon direction score
0.39
0.03
0.43
0.12
0.48
0.11
0.51
0.23
0.58
0.19
1965-1974
1975-1984
1985-1994
1995-2004
2005-2016