Jonathan D. Moreno

David and Lyn Silfen
University Professor

Penn
UNIVERSITY of PENNSYLVANIA

Penn Symposium on Social Implications of Autonomous Military Systems

May 24, 2021

CHINA

MAY 05, 2021 / 11:42 AM

Top U.S. general warns of 'potential international instability'

By Reuters Staff

Army General Mark Milley, chairman of the Joint Chiefs of Staff, compared the current era to other major geopolitical shifts in world history, including the fall of Rome and the collapse of the Soviet Union.

He warned that the geopolitical change was being accompanied by technological innovation in robotics, hypersonic weapons, artificial intelligence and other technologies.

Here's my argument

Over the last two decades or so the debate has moved through several overlapping phases:

- the ethics of "autonomous" robotic weapons

- excitement about autonomous weapons systems

- trying to figure out what autonomy means

- pulling back on the enthusiasm

- a shift to "teaming" or "collaboration"

Note: AI seems to be inextricable from autonomy but we don't know what AI means either.

MIND WARS

BRAIN SCIENCE AND THE MILITARY IN THE 21ST CENTURY

EMERGING COGNITIVE NEUROSCIENCE
AND RELATED TECHNOLOGIES

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

Roots: Ron Arkin and his critics

NewScientist

News Podcasts Video Technology Space Physics Health More ⌄ Shop Courses Events

Plan to teach military robots the rules of war

TECHNOLOGY 18 June 2009

By Tom Simonite

Moral Machines
Teaching Robots Right from Wrong

Wendell Wallach · Colin Allen

Governing Lethal Behavior in Autonomous Robots

Ronald C. Arkin

CRC Press
A CHAPMAN & HALL BOOK

ROBOT ETHICS
THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS

EDITED BY
Patrick Lin, Keith Abney,
and George A. Bekey

"Autonomous armed robotic platforms may…better adhere to the laws of war than most soldiers possibly can."
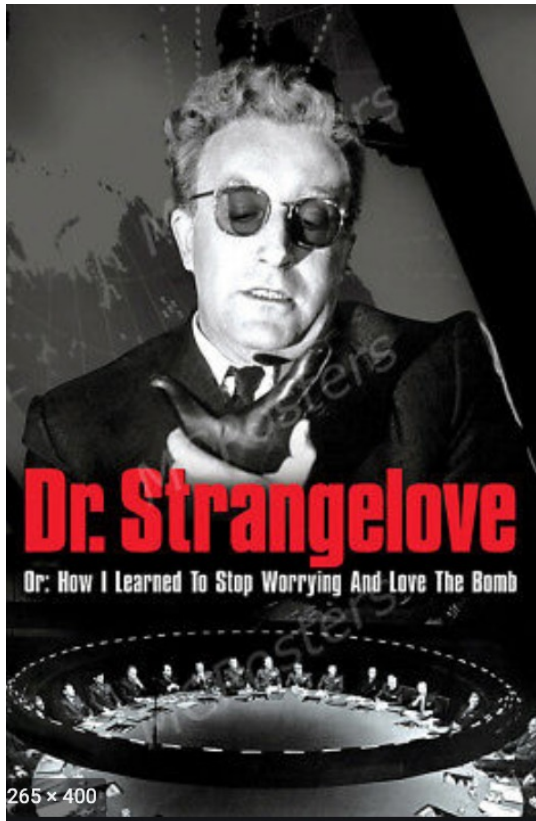
Ron Arkin, "Ethical Robots in Warfare," 2009
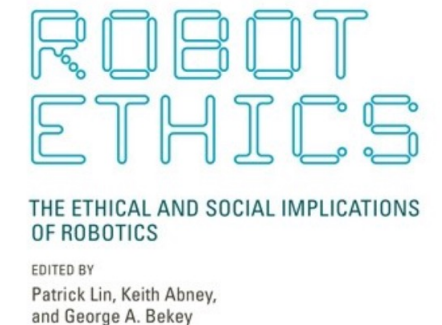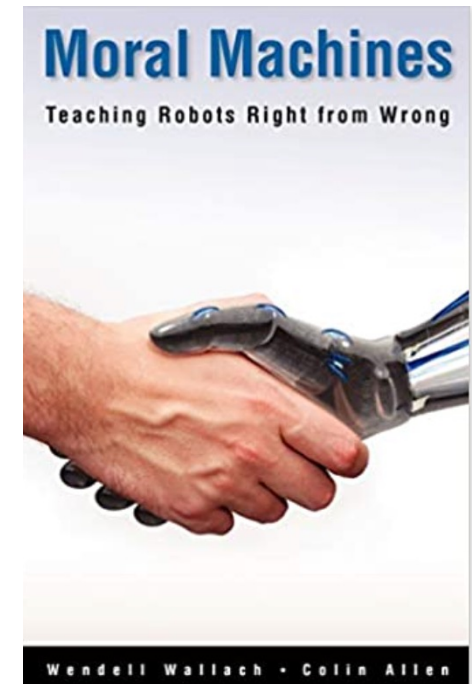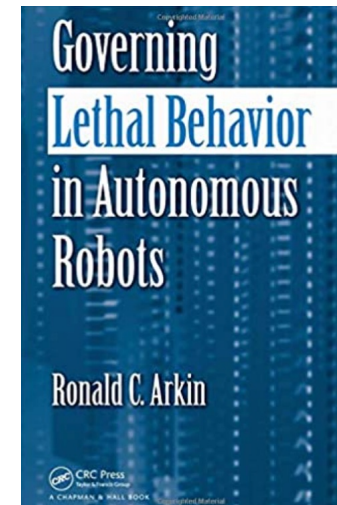
Automatic vs. Autonomous

The doomsday machine is automatic:
Highly restricted parameters and tasks.

Skynet is autonomous: *learns and adapts to dynamic environments, evolves as the environment around it changes, ingests and learns from increasing data sets faster, and eventually more reliably, than what would be reasonable for a human.* (definitions paraphrased from Scott Totman, 2017)

"If Stanley Kubrick directed 'Dr. Strangelove again, it would be about the issue of autonomous weapons."

-Michael Schrage, MIT

# It's not going to end with robots fighting robots.

Basic premises

- LAWS must follow the principles of just war theory: sovereign authority, just cause, just intentions, peace as goal, proportionality, distinction, etc.

- Lethality cannot always be reliably predicted (e.g., Moscow theater hostage crisis, 2002).

- Autonomous weapons systems are *emerging* technologies, thus inherently difficult to regulate (Collingridge Dilemma, 1980).

- The first mover must set the rules. That almost never happens.

Autonomous technologies are "dual use"

Civilian Benign	Military Benign
Military Malign	Civilian Malign

A single technology may originally fit only one cell (e.g, Skynet), in a couple of cells (e.g., robotic surgery), or every cell (e.g., surveillance) depending on circumstances.
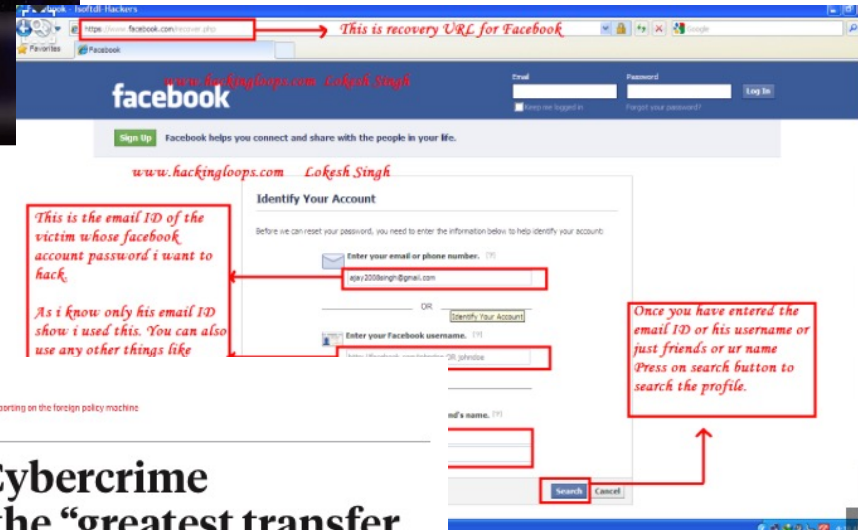
Distinction between "military" and "civilian" AI increasingly irrelevant (Allenby)

"Autonomy" is a many splendored thing

- Political philosophy: self-governance

- Kantian moral philosophy: ability to self-legislate the moral law

- Robotics: capacity of a system to make its own "decisions"*

  - "Consciousness" not required except perhaps in the cognitive sense of systematic information processing

*"A metaphor is an affair between a predicate with a past and an object that yields while protesting."

--Nelson Goodman

Unmanned Military Machines: "In the loop"

- US Civil War—Balloons to deliver explosives

- Tesla – Wirelessly controlled motorboats (1915)

- WWII – German FX-1400 remote aircraft

- WWII – US B-17s remote controlled

- Cold War – UAVs for images and heavy payloads

Weapons Systems Classified by Relative Autonomy

- Automatic weapons ("dumb" -- e.g. triggered by a tripwire or motion detector; the doomsday device)

- GOFAI-based weapons (human-readable

  representations of problems, logic and searches;

  e.g., PackBot)

- Computational AI-based weapons (e.g., machine learning, neural net architecture)
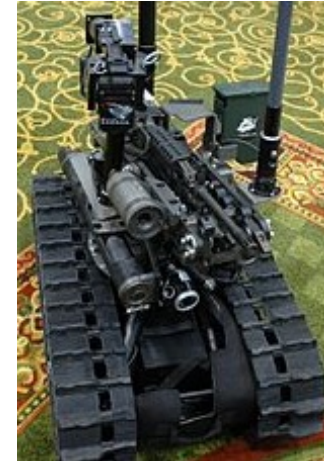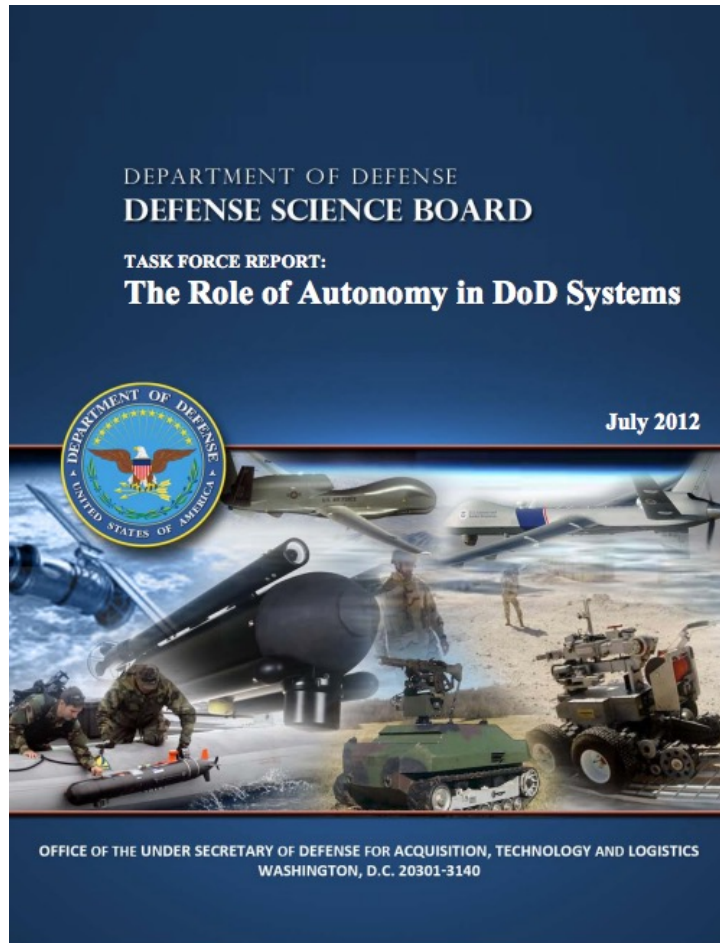
Increasing Control Autonomy: UAVs "On the loop"

- Gulf War – RQ-2 Pioneer for surveillance

- Afghanistan – MQ-1 Predator, RQ-1 Global Hawk for armed surveillance

Today: Continuous military demand for self-directing, goal oriented, environmentally adaptive and rapid "sense and trigger" scenarios.

2012

Department of Defense
DEFENSE SCIENCE BOARD

TASK FORCE REPORT:
The Role of Autonomy in DoD Systems

July 2012

OFFICE OF THE UNDER SECRETARY OF DEFENSE FOR ACQUISITION, TECHNOLOGY AND LOGISTICS
WASHINGTON, D.C. 20301-3140

THE WALL STREET JOURNAL.

SUBSCRIBE NOW and get 3 MONTHS for the PRICE OF 1 — SUBSCRIBE NOW

U.S. EDITION ▾ Friday, May 11, 2012 As of 6:07 PM EDT

Home | World | U.S. | New York | Business | Tech | Markets | Market Data | Opinion | Life & Culture | Real Estate | Management | C-Suite

Editorials | Features | Columns | Political Diary | Peggy Noonan's Blog | Book Reviews | Leisure & Arts | Letters to the Editor

TOP STORIES IN OPINION
1 of 12 — The President's Priorities
2 of 12 — Henninger: How the East Was Won
3 of 12 — Karl Rove: Obama Gears Up to 'Flip the House' in 2014
The Terror Merger

OPINION | May 11, 2012, 6:07 p.m. ET

Robot Soldiers Will Be a Reality—and a Threat

Given the obvious dangers, fully autonomous offensive lethal weapons should never be permitted.

Article | Comments (35)   MORE IN OPINION ▸

Email | Print | Save   f t g+ in   A A

The world doesn't need another luxury car.

By JONATHAN D. MORENO

Much controversy has surrounded the use of remote-controlled drone aircraft or "unmanned aerial vehicles" in the war on terror. But another, still more awe-inducing possibility has emerged: taking human beings out of the decision loop altogether.

CAMPAIGN TO STOP KILLER ROBOTS

Brief on Autonomy Initiatives in the US DoD

8 November 2012

Autonomy Priority Steering Council

Key DOD Challenges Addressed by Autonomy

Manpower efficiencies: Insufficient manpower to support complex missions such as command and control and surveillance across relevant battlespace

Harsh environments: Operational environments that do not reasonably permit humans to enter and sustain activity

New mission requirements: Need adaptive autonomous control of vehicle systems in face of unpredictable environments and challenging missions

What is "Automated" Technology?

Automation: Using machines to accomplish tasks traditionally performed by humans. Automated systems are most effective in predictable environments. Automation is not limited to simple tasks, but rather to well defined tasks with predetermined responses to all operational contingencies.

Historic Challenges Associated with Automated Technology:

Technical	Computer processing speed
	Sensor development/integration
	Cyber/mechanical teaming
Social	Human trust in automation
	Impact upon work force/political
	Human-machine teaming
Economic	Significant initial investment/maintenance

The Automation-Autonomy Continuum

Automation

Autonomy

Primary sources of automation brittleness:

1. Dynamic and complex mission *requirements*

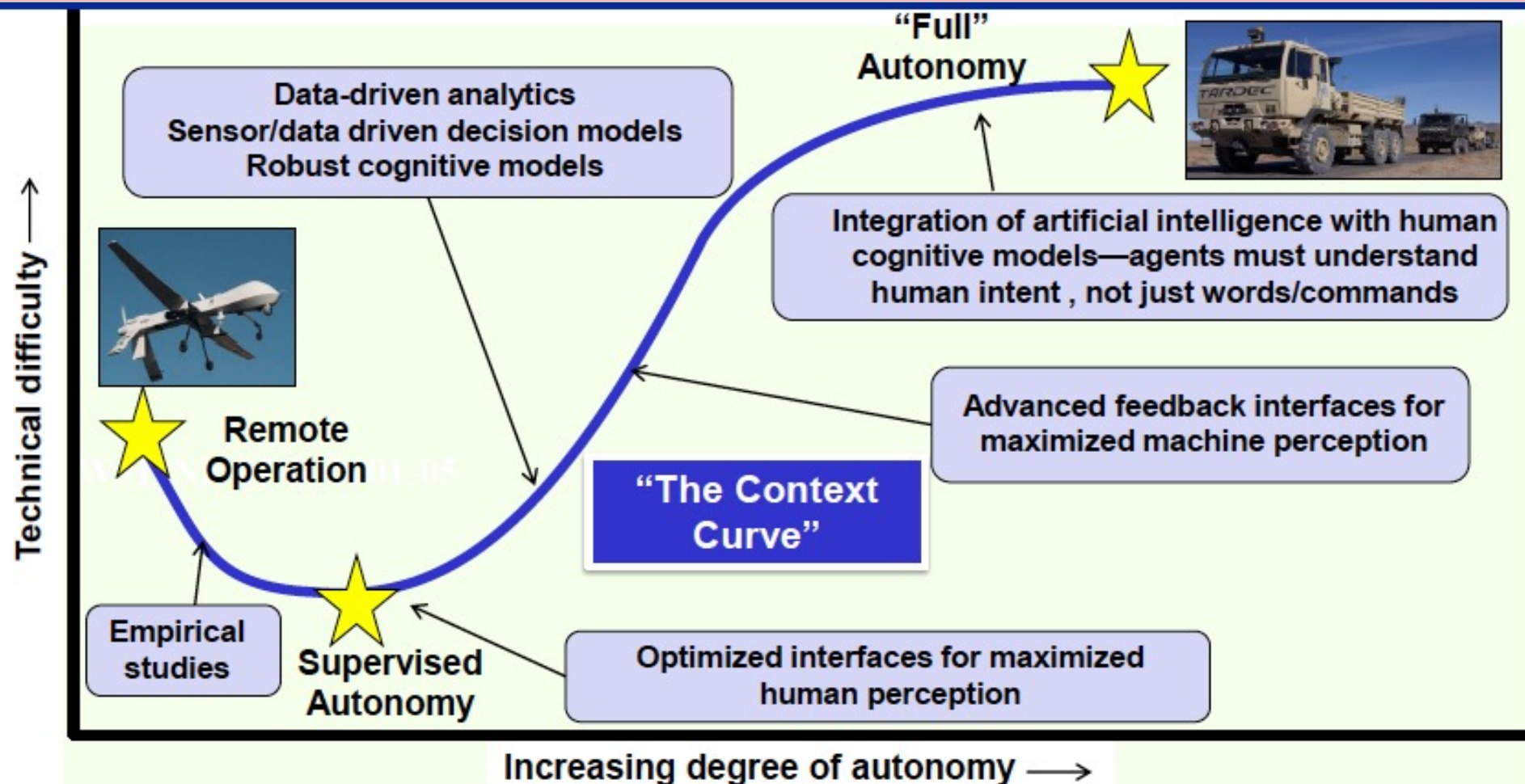2. Dynamic and complex operational *environments*
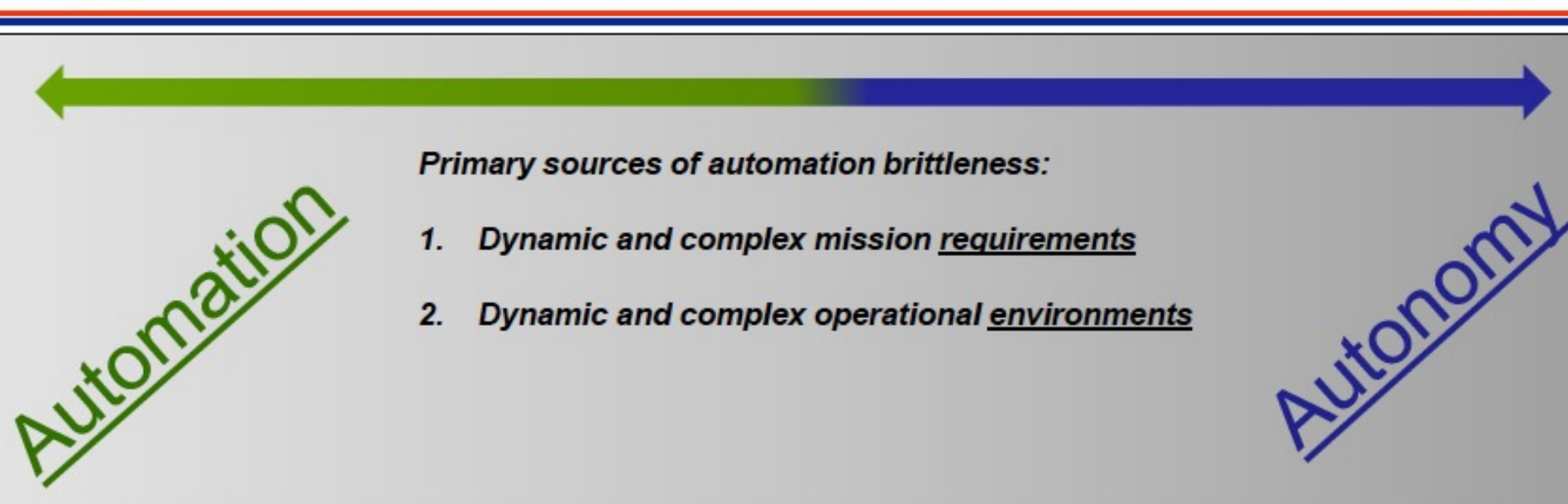
In a static environment, with a static mission, automation and autonomy converge. However, in reality, where dynamic environments collide with dynamic missions, automation can only support a small fraction of autonomy requirements.

Autonomy Capability Curve

"Full" Autonomy

Data-driven analytics
Sensor/data driven decision models
Robust cognitive models

Integration of artificial intelligence with human cognitive models—agents must understand human intent, not just words/commands

Advanced feedback interfaces for maximized machine perception

Technical difficulty →

Remote Operation

"The Context Curve"

Empirical studies

Supervised Autonomy

Optimized interfaces for maximized human perception

Increasing degree of autonomy →

👤 / ✈ > 1   →   👤 / ✈ < 1

Data comprehension (that of the human and his agent) drives functionality

The U.S. "Third Offset" Strategy
(SecDef Chuck Hagel, 2014)

- Robotics

- *Systems autonomy*

- Miniaturization

- Big data

- Advanced manufacturing

Partner with innovative private sector companies

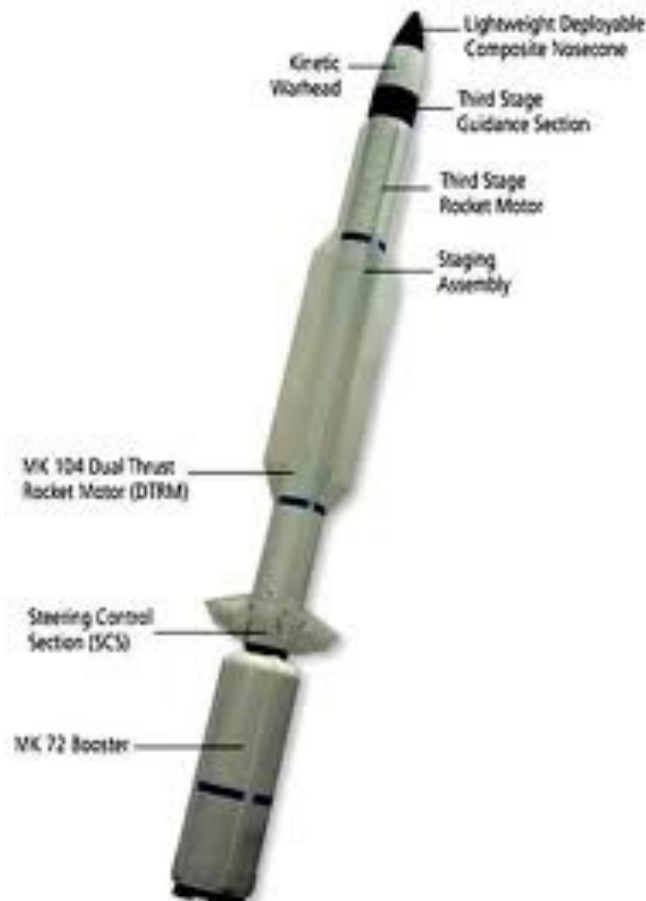# Future of Life Institute, 2015:
"Meaningful human control"
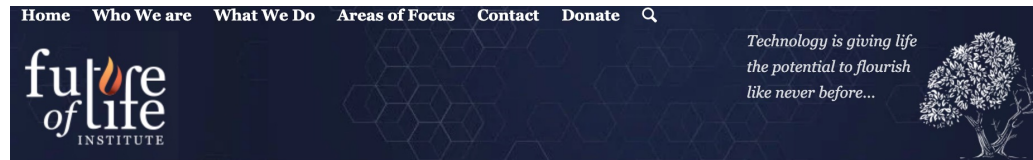
AUTONOMOUS WEAPONS: AN OPEN LETTER
FROM AI & ROBOTICS RESEARCHERS

In summary, we believe that AI has great potential to benefit humanity in many ways, and that the goal of the field should be to do so. Starting a military AI arms race is a bad idea, and should be prevented by a ban on offensive autonomous weapons beyond meaningful human control.

"Meaningful human control"?

Kinetic Warhead

Lightweight Deployable Composite Nosecone

Third Stage Guidance Section

Third Stage Rocket Motor

Staging Assembly

MK 104 Dual Thrust Rocket Motor (DTRM)

Steering Control Section (SCS)

MK 72 Booster

The Washington Post — FINAL

Navy Missile Downs Iranian Jetliner Over Gulf; Iran Says 290 Are Dead; U.S. Expresses Regret

Pentagon Says Radar on Aegis Cruiser Mistook Airbus for Attacking F14

"Meaningful human control"?

"The Tomahawk is already an autonomous weapon, in the sense that it can be fired at pre-planned fixed targets and left alone."
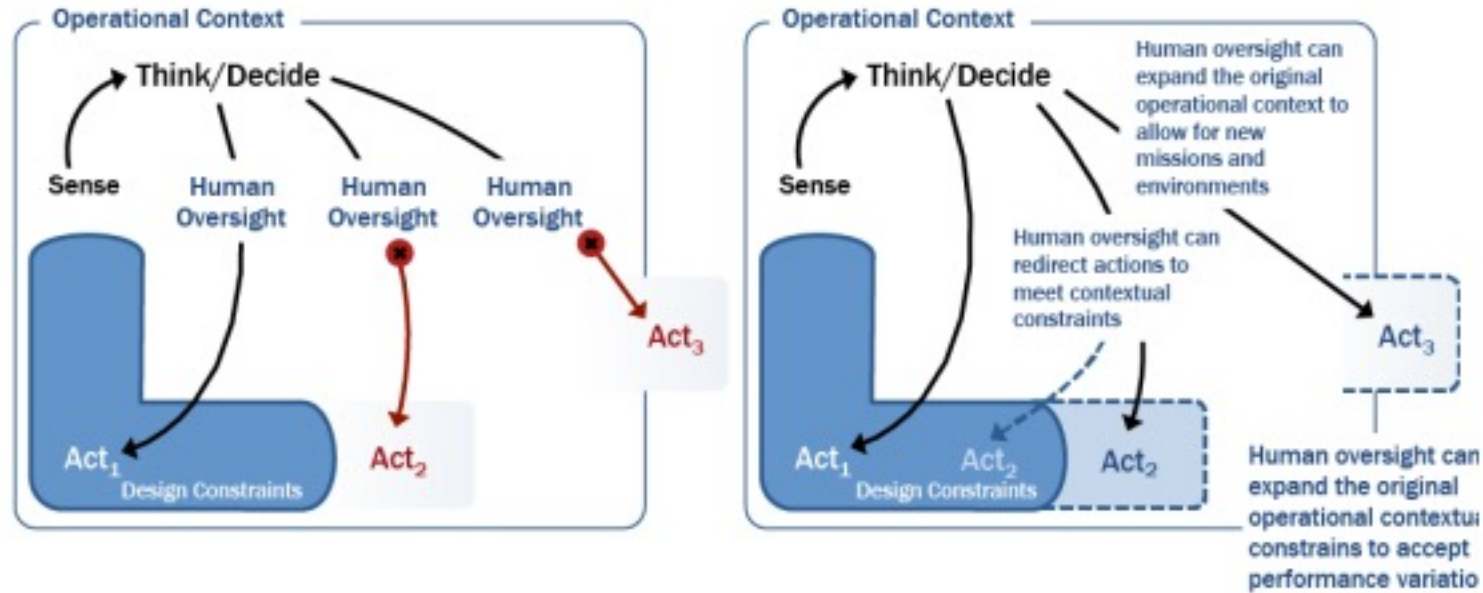
--Defense Industry Daily, 12/8/20

Defense Science Board

SUMMER STUDY ON AUTONOMY

JUNE 2016

Figure 6   Oversight "on-the-loop" provides additional opportunities for human-machine partnership.

Pressure on the loop

"In short, *speed matters* – in two distinct dimensions. First, autonomy can increase decision speed, enabling the U.S. to act inside an adversary's operations cycle. Secondly, ongoing rapid transition of autonomy into warfighting capabilities is vital if the U.S. is to sustain military advantage."
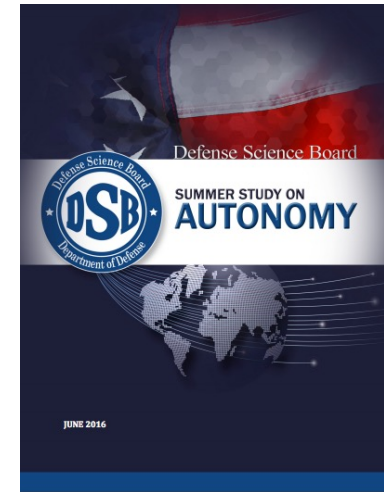
- Autonomous systems "must have the capability to independently compose and select among different courses of action to accomplish goals based on its knowledge and understanding of the world, itself, and the situation."

- The U.S. should take "immediate action" to figure out how to defeat new AI-enabled operations.
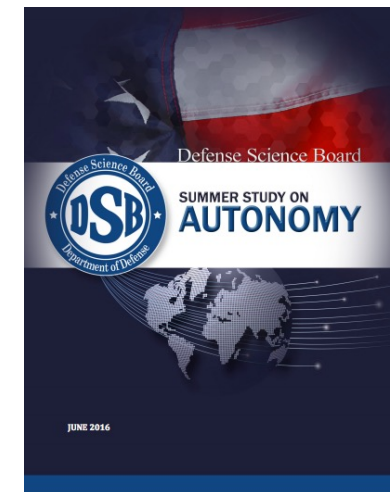
Correcting "Misperceptions"

- All autonomous systems are "*supervised* by human operators at some level."

- Design and operation involve "*human-system collaboration*."

- Commanders and operators may mistrust software.

- *DoD "should abandon the debate*

  over definitions of levels of autonomy."
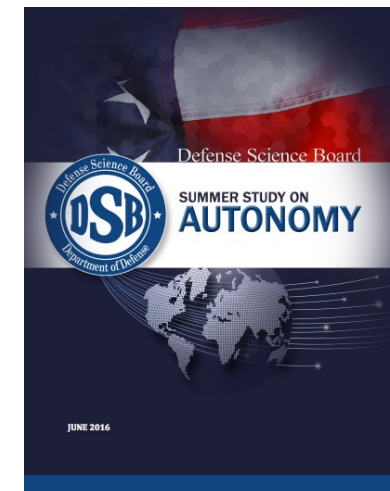
"One of the places where we spend a great deal of time is determining whether or not the tools we are developing absolve humans of the decision to inflict violence on the enemy. That is a fairly bright line that we're not willing to cross."

General Selva, August 25, 2016

WASHINGTON — As autonomous technology continues to evolve, the Pentagon finds itself being pulled in two directions, enticed by the capabilities that autonomous systems could provide while also insistent it always be subservient to humans, and a set of human morals and mindsets.

That tension was on full display Aug. 25, when a new report from a key Pentagon advisory group called for an acceleration of autonomous systems within the US military at the same time the country's second highest ranking uniformed officer warned that there will need to be limits on how the technology is used in order to avoid the dreaded killer-robot scenario.
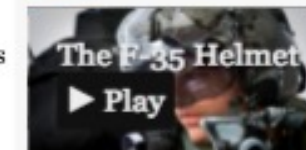
"A raucous debate…"

CNN politics | 45 CONGRESS SECURITY THE NINE TRUMPMERICA STATE

U.S. MILITARY

US general warns of out-of-control killer robots

By Ryan Browne, CNN
Updated 9:18 PM ET, Tue July 18, 2017
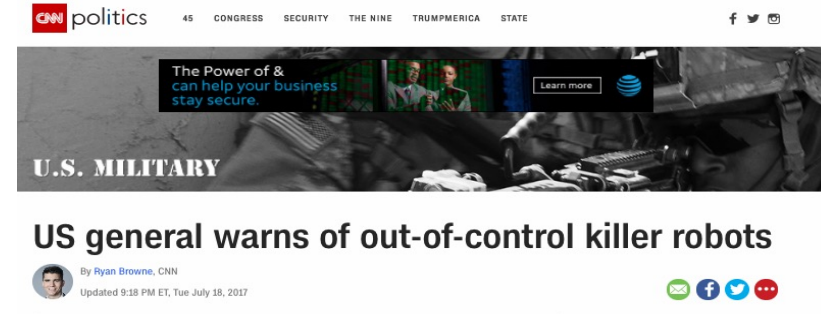
"I must tell you I'm an advocate for keeping that restriction because we take our values to war, and because many of the things that we must do in war are governed by the laws of war, which say we must take proportional and discriminate action to action against our enemy to achieve our objectives.

General Paul Selva, Joint Chiefs Vice Chair
Senate testimony, July 18, 2017

"People say 'what's the Third Offset Strategy about? And they say 'oh, it's about AI and autonomy.' We say no...It's about human-machine collaborative combat networks."

Interview with Deputy Defense Secretary Robert Work, *Breaking Defense*, February 9, 2016

DoD Directive 3000.09 on "Autonomy in Weapon Systems"

1) Semi-autonomous weapon systems (including manned or unmanned platforms, munitions, or sub-munitions that function as semi-autonomous weapon systems or as subcomponents of semi-autonomous weapon systems) may be used to apply lethal or non-lethal, kinetic or non-kinetic force. **Semi-autonomous weapon systems** that are onboard or integrated with unmanned platforms must be designed such that, in the event of degraded or lost communications, the system **does not autonomously select and engage individual targets** or specific target groups that have not been previously selected by an authorized human operator.

2) Human-supervised autonomous weapon systems may be used to select and engage targets, with the exception of selecting humans as targets, for local defense to intercept attempted time-critical or saturation attacks for:

   (a) Static defense of manned installations.

   (b) Onboard defense of manned platforms

Future Armament Systems will require "Supervised Autonomy" Capabilities

TAKEAWAY

- **ARDEC has demonstrated lethality on a number of UGVs**

- **Future Armament Systems <u>will</u> require autonomous and remote operation capabilities**

- **Must address Supervised Autonomy holistically at the Integrated Armament System level**

- **'Ground Floor' Armament Architecture Development Required.**

Innovations

The Pentagon is building robotic wingmen to fly alongside fighter planes

By Aaron Gregg June 14 ✉

A futuristic rendering of what Kratos' semiautonomous drones might look like in combat (Courtesy of Kratos Defense and Security Solutions)

A Pentagon effort to incubate businesses in Silicon Valley may be bearing some of its first fruit, as a San Diego company rolls out a set of new drones it says could accompany human-piloted fighters into combat.

On Tuesday, Kratos Defense and Security Solutions officially announced two new classes of drones designed to function as robotic wingmen for fighter pilots. Development of the UTAP-22 Mako has been funded by the Defense Department's Silicon Valley laboratory, dubbed DIUx. Separately, the company showed off a larger, 30-foot-long drone backed by the Air Force called the XQ-222 Valkyrie, with a range of more than 4,000 nautical miles. Kratos is promoting the pilotless planes at the Paris Air Show next week in preparation for a new round of testing.

DOD Adopts Ethical Principles for Artificial Intelligence

FEB. 24, 2020

1. Responsible. DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.

2. Equitable. The Department will take deliberate steps to minimize unintended bias in AI capabilities.

3. Traceable. The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation.

4. Reliable. The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles.

5. Governable. The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.

AIR WARFARE, ALLIES, NETWORKS / CYBER, PENTAGON

## AI To Fly In Dogfight Tests By 2024: SecDef

- Sydney J. Freedberg Jr.

5 months ago

Mark Esper checks out an Army helicopter cockpit during a visit to Fort Knox, Ky.

WASHINGTON: Defense Secretary Mark Esper used the Pentagon first AI conference to issue a challenge to China and Russia. The US, he vowed, will lead the world on the military use of artificial intelligence – including testing an AI pilot in a fighter by 2024. But, he said, America's AI will be governed by ethics that its great power rivals lack, and it will be coordinated with nearly a dozen democratic allies in a new "AI Partnership for Defense."
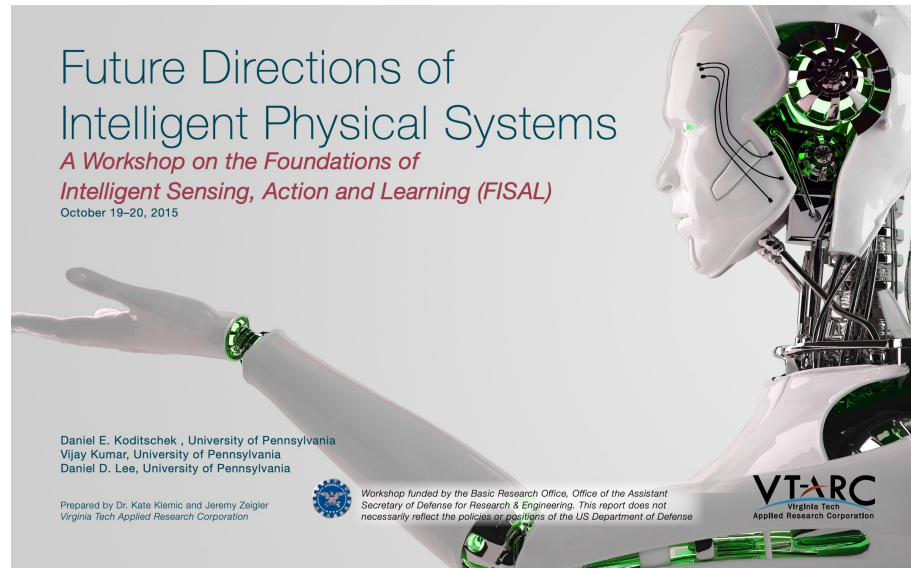
Social interaction – developing new theories of intent, trust and collaboration that lie at the heart of the physically mediated social dynamics necessary to field a human-machine team.

ARIN BURGESS / MILITARY REVIEW

IDEAS

Will Commanders Trust Their New AI Weapons and Tools?

Example of teaming technology: N3

DARPA Funding for BCI over the decades

Most programs are in the USD 50-100 Million USD range, and overall funding for invasive interfaces has been higher than non-invasive ones
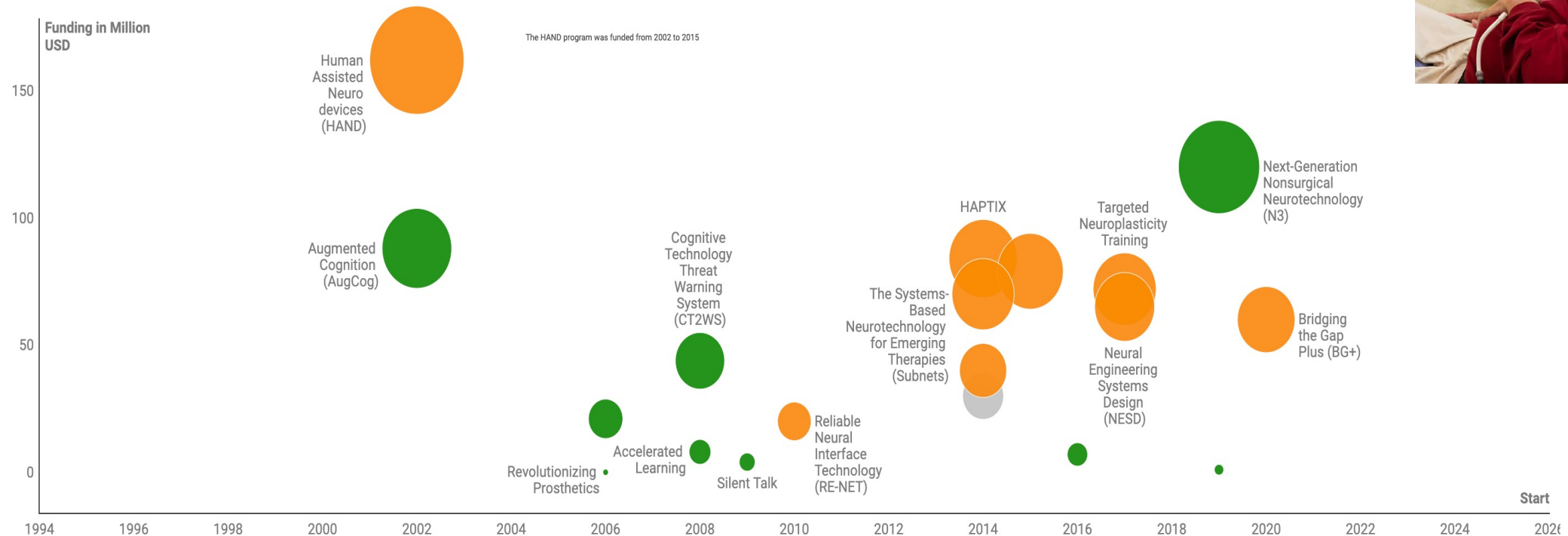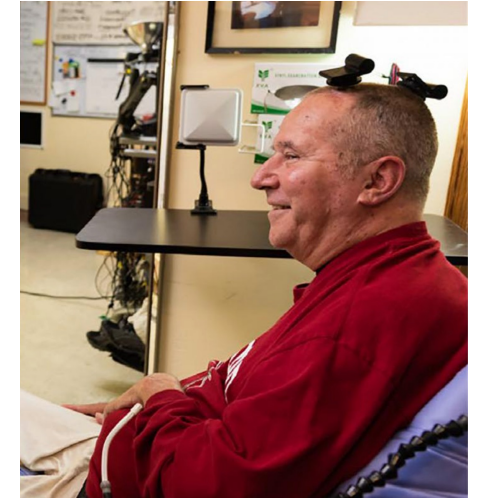
● Invasive ● Non-invasive ● Other

Funding in Million USD

The HAND program was funded from 2002 to 2015

Human Assisted Neuro devices (HAND)

Augmented Cognition (AugCog)

Cognitive Technology Threat Warning System (CT2WS)

HAPTIX

Targeted Neuroplasticity Training

Next-Generation Nonsurgical Neurotechnology (N3)

The Systems-Based Neurotechnology for Emerging Therapies (Subnets)

Neural Engineering Systems Design (NESD)

Bridging the Gap Plus (BG+)

Revolutionizing Prosthetics

Accelerated Learning

Silent Talk

Reliable Neural Interface Technology (RE-NET)

Start

150
100
50
0

1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020 2022 2024 2026

Chart: Pooja Rao • Source: News reports and DARPA's website • Created with Datawrapper

The Ethics of Warfighter Participation in the Development and Testing of AI-Driven Performance Enhancements

PIs: Nick Evans and Neil Shortland (UMass Lowell)

Investigators: Michael Gross (Haifa) and Jonathan D. Moreno (Penn)

Human Experiments in the Military

- Because **investigational drugs** offer a high probability of protecting fighting forces, military necessity can override patients' rights **by presidential waiver** in the US when obtaining consent is not feasible.

- **Experimental drugs**, on the other hand, **are risky and not necessarily intended to protect the health of the test subject**. Regulations, therefore, are stricter. Military subjects, moreover, are particularly vulnerable and although soldiers sign **consent forms**, problems arise because the dynamics of **rank disparity**, **fears of offending one's superiors and/or peer pressure** may undermine informed consent when soldiers are recruited for medical experiments.

- US Department of Defense (DoD) military human research protection program (HRPP) **forbids the presence of superior officers** during the solicitation of research subjects and demands **informed consent, medical supervision, the right to end an experiment, an independent ombudsman or research monitor** to oversee recruitment and experimentation.

Specific Aims of Our Project

What ethical considerations inform the use of warfighters as experimental subjects in neuroscience research, and as users of experimental neurotechnologies?

- Identify special ethical features of warfighters as experimental subjects as they pertain to the testing and use of neuroenhancements for restoring and/or maintaining human performance and capacity;

- Investigate how military commanders weigh the risks and benefits of experimental technologies used by or on their warfighters.

Significance

Within the study of ethics and neuroscience, little research has been devoted to explicit concerns at the **intersection of neuroscience and national security**. A secondary literature on enhancement has focused on the use of neuroscience to augment warfighter capacities beyond therapeutic norms through enhanced training or the use of brain-computer interfaces to connect soldiers directly to data analysis or prosthetic devices.

To date, however, there has not been an **explicit investigation of the ethics of testing and implementing enhancements,** with a particular emphasis on so-called "neuroenhancements" (i.e. neuroscience and technology used to enhance capacities), in warfighters.

Innovation

There is a **lack of sustained analysis of what ethical issues arise when interventions are not clearly or centrally therapeutic, but are rather to enhance warfighter capability.** The scholarly intersection of military medical ethics, neuroscience, and enhancement, moreover, is almost totally empty. There has been almost no specific analysis about how warfighter enthusiasm for enhancement could be exploited by researchers to test novel neuroscience interventions.

Approach

We conduct a **historical inquiry** into the role of soldiers in testing novel medical interventions and technologies.

We conduct a **normative inquiry** into the ethical foundations of human subjects research and how they apply in the context of warfighters as a distinct population.

Research Design

We engage in a **historical and normative analysis** of the ethics of experimentation on warfighters when the intervention to be studied constitutes an enhancement.

Our methodology includes **an expert workshop and a horizon scan** of stakeholders: planners, logisticians, SOF operators, defense AI specialists, military medical researchers.

Our Team
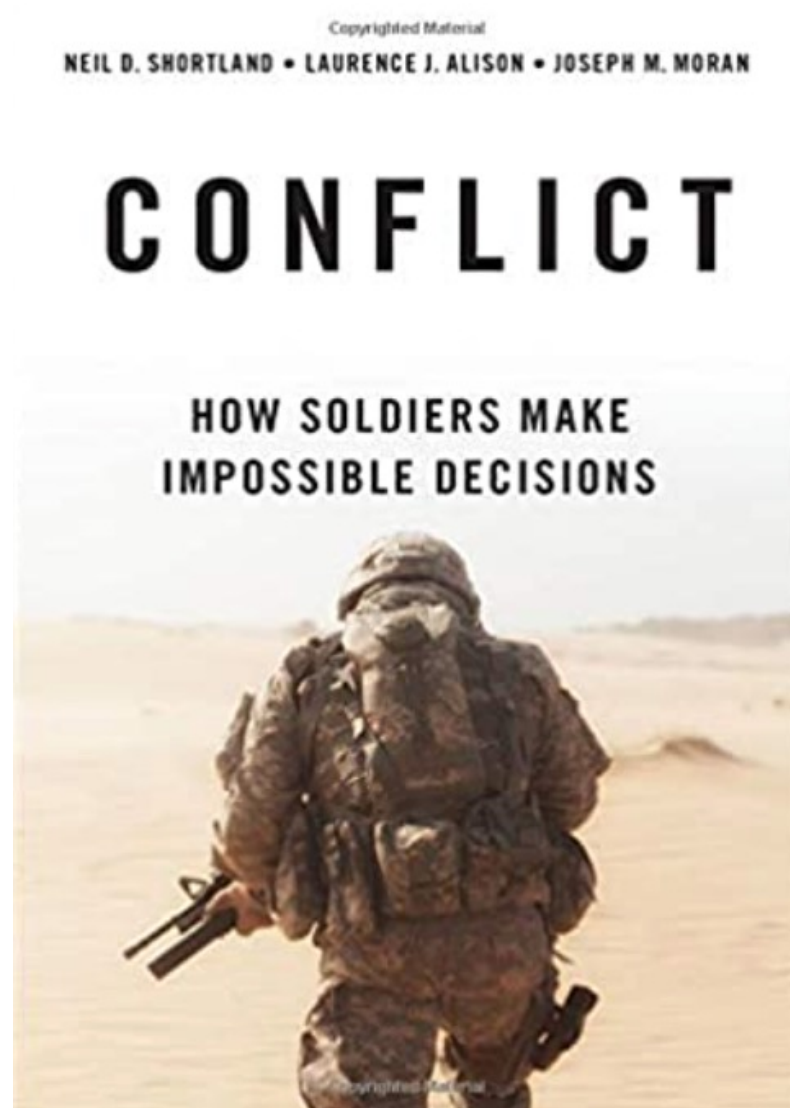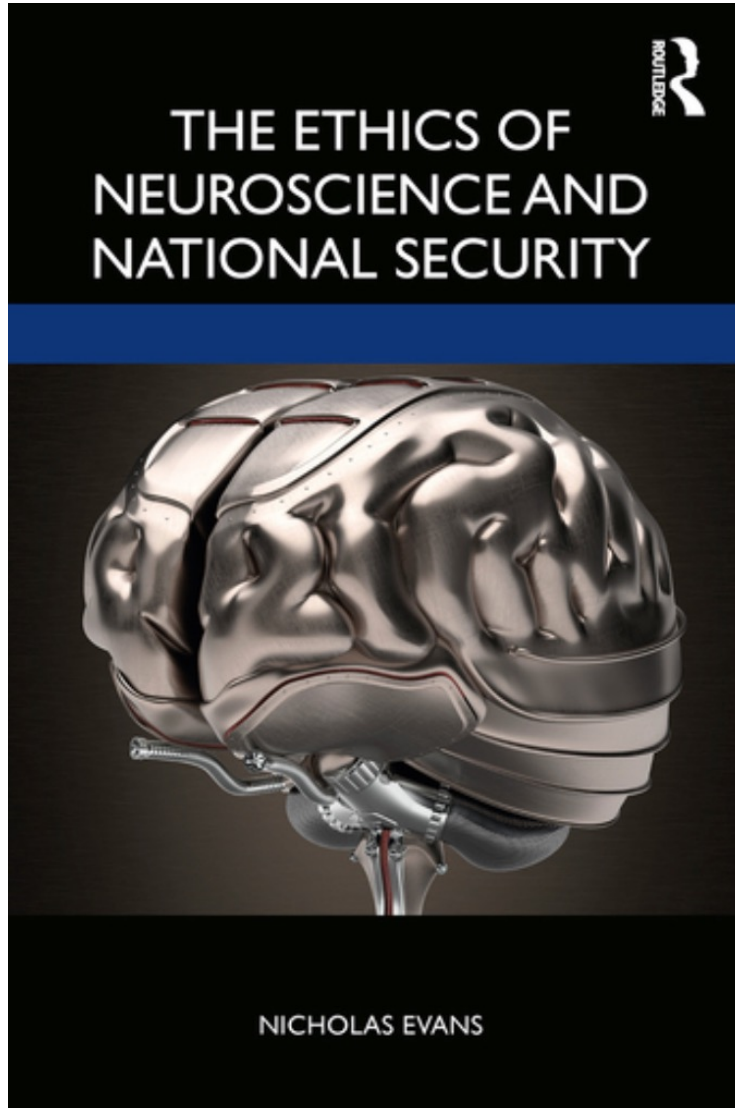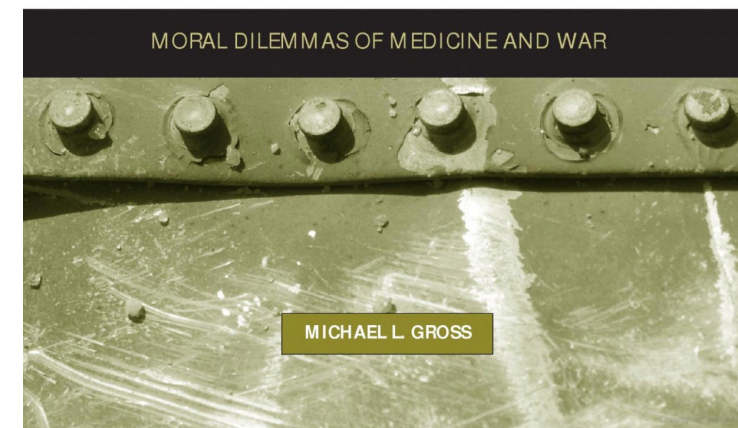
A Parting Shot….

"Despite the fact that AI has not been as successful in military and commercial settings as many people think, it is entirely possible that the perception of having all-powerful AI may be just as important as actually having it."

--Missy Cummings, Duke University

Thank you!