

THE “BRIGHT GREEN LINE” OF RESPONSIBILITY

Mark R. Waser
Digital Wisdom Institute
MWaser@DigitalWisdomInstitute.org

IN DEFENSE OF SMART MACHINES

Mark R. Waser
Digital Wisdom Institute
MWaser@DigitalWisdomInstitute.org

The Killer Robot War Is Coming

The new laws we need to govern the use of drones.

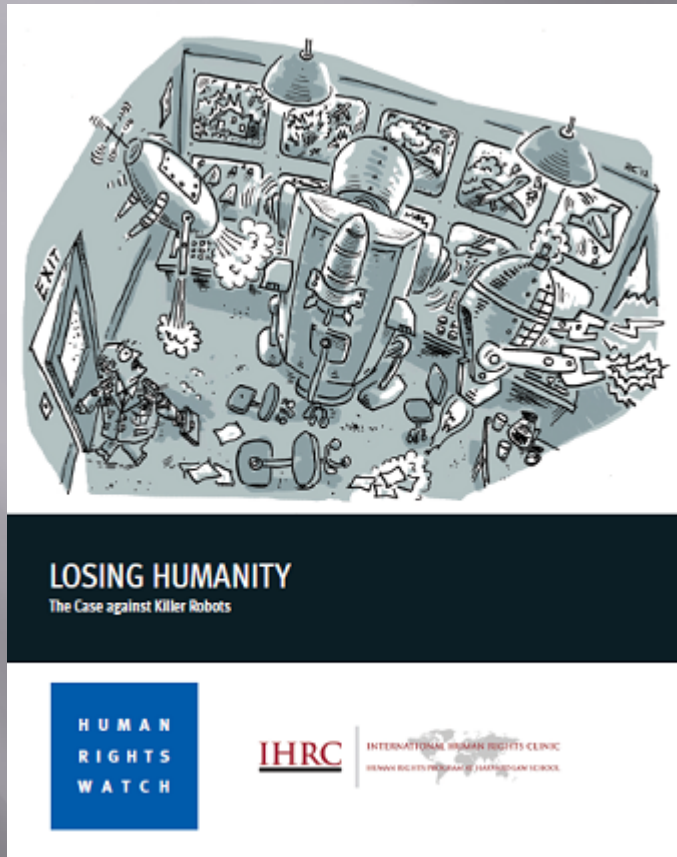
By **Eric Posner** | Posted Wednesday, May 15, 2013, at 2:57 PM



An X-47B Unmanned Combat Air System demonstrator flies over the flight deck of the aircraft carrier USS George H.W. Bush (CVN 77) on May 14, 2013, in the Atlantic Ocean.

Photo by Mass Communication Specialist 2nd Class Timothy Walter/U.S. Navy via Getty Images

The Killer Robot War Is Coming . . .



November 2012

Given the rapid pace of development of military robotics and the pressing dangers that these pose to peace and international security and to civilians in war, we call upon the international community to urgently commence a discussion about an arms control regime to reduce the threat posed by these systems.

We propose that this discussion should consider the following:

Their potential to lower the threshold of armed conflict;

The prohibition of the development, deployment and use of armed autonomous unmanned systems; machines should not be allowed to make the decision to kill people;

Limitations on the range and weapons carried by "man in the loop" unmanned systems and on their deployment in postures threatening to other states;

A ban on arming unmanned systems with nuclear weapons;

The prohibition of the development, deployment and use of robot space weapons. —Mission Statement

ICRAC

International Committee for Robot Arms Control

Home Who We Are Statements Resources Supporters' Network The Scientists' Call Contact Private + Subscribe to RSS

The Scientists' Call

... To Ban Autonomous Lethal Robots

As Computer Scientists, Engineers, Artificial Intelligence experts, Roboticians and professionals from related disciplines, we call for a ban on the development and deployment of weapon systems in which the decision to apply violent force is made autonomously.

We are concerned about the potential of robots to undermine human responsibility in decisions to use force, and to obscure accountability for the consequences. There is already a strong international consensus that not all weapons are acceptable, as illustrated by wide adherence to the prohibitions on biological and chemical weapons as well as anti-personnel land mines. We hold that fully autonomous robots that can trigger or direct weapons fire without a human effectively in the decision loop are similarly unacceptable.

Demands within the military for increasingly rapid response times and resilience against communications failures, combined with ongoing investments in automated systems, indicate a trend towards fully autonomous robotic weapons. However, in the absence of clear scientific evidence that robot weapons have, or are likely to have in the foreseeable future, the functionality required for accurate target identification, situational awareness or decisions regarding the proportional use of force, we question whether they could meet the strict legal requirements for the use of force. This is especially true under conditions in which battlefields are not clearly delimited and discrimination between civilians, insurgents and combatants is increasingly difficult.

Moreover, the proliferation of autonomous robot weapons raises the question of how devices controlled by complex algorithms will interact. Such interactions could create unstable and unpredictable behavior, behavior that could initiate or escalate conflicts, or cause unjustifiable harm to civilian populations.

Given the limitations and unknown future risks of autonomous robot weapons technology, we call for a prohibition on their development and deployment. Decisions about the application of violent force must not be delegated to machines.

ICRAC SUPPORTS CSKR

CAMPAIGN TO STOP KILLER ROBOTS

SEARCH

CATEGORIES

- Analysis
- ICRAC in the media
- ICRAC News
- News
- Radio Programs
- Television Programs
- YouTube videos

RECENT POSTS

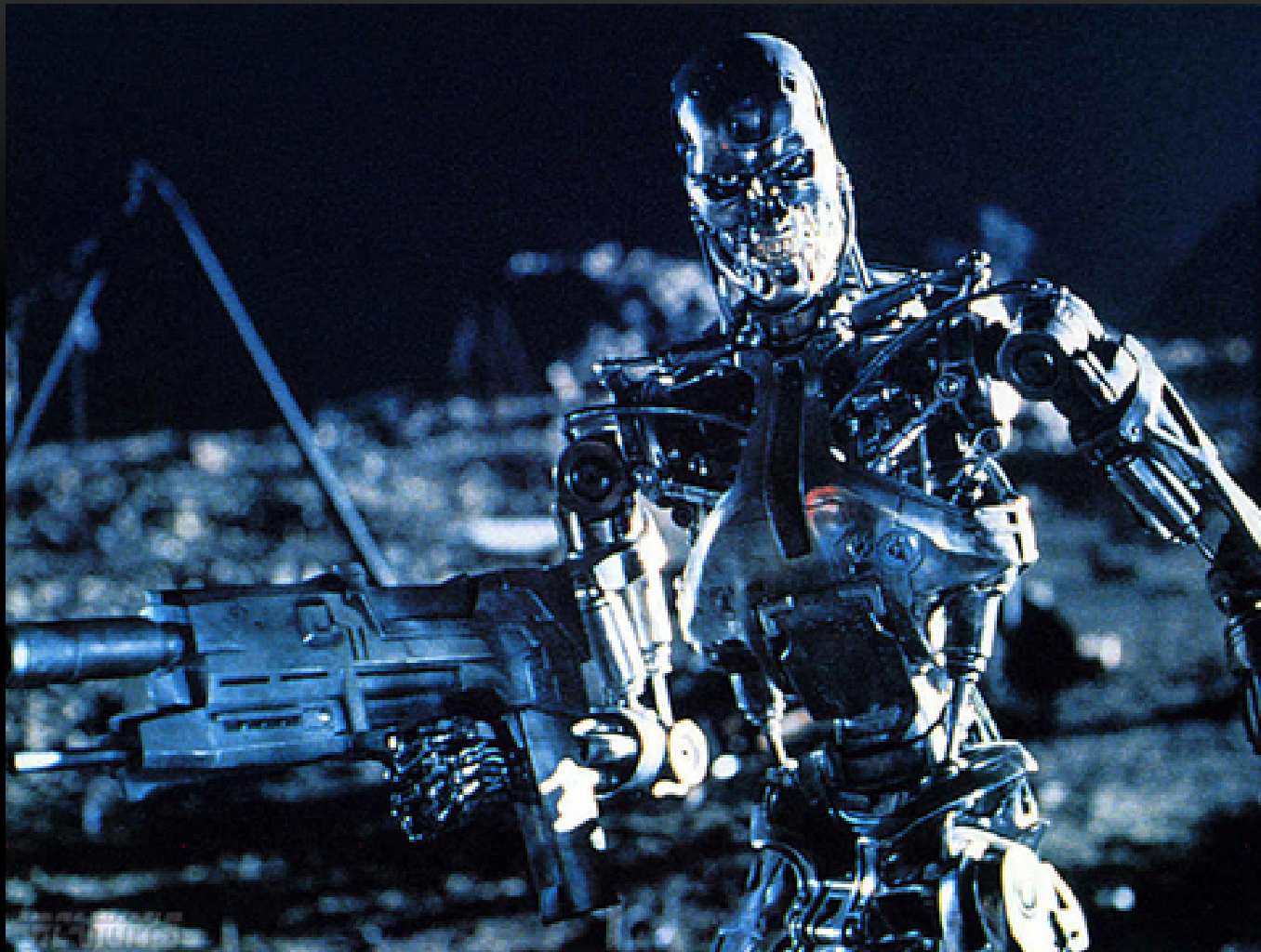
- Engineers strongly favour a total ban on killer robots
- The Role of ICRAC in the Arms Trade Treaty Negotiations
- Arms Control for Uninhabited Vehicles: A Detailed Study

TWITTER

- ICRACnet about 6 days ago Watch a summary of the press conference of the Campaign to Stop #KillerRobots in London @bankillerbots: youAa be/PQ70G0VrR
- ICRACnet about 6 days ago Impressions from the first IVO conference of the Campaign To Stop #KillerRobots in London @bankillerbots: youAa be/IS1KJ7FUz
- ICRACnet about 15 days ago New posting: Engineers strongly favour a total ban on killer robots - tinyurl.com/d0u3na
- ICRACnet about 15 days ago "The Point of No Return": Should Robots Be Able to Decide to Kill You On Their Own? rol at/12Qqj51 via @rollingstone
- ICRACnet about 17 days ago A brief overview over the Campaign launch in London

February 2013

April 2013 – UN GA/HRC - Lethal autonomous robotics and the protection of life



Risk of a Terminator Style Robot Uprising to be Studied

Technology Governance Curves

ICRAC - The Scientists' Call

... To **Ban** Autonomous Lethal Robots

As Computer Scientists, Engineers, Artificial Intelligence experts, Roboticists and professionals from related disciplines, we call for a **ban** on the development and deployment of weapon systems in which the decision to apply violent force is made autonomously.

Decisions about the application of violent force **must not** be delegated to machines.

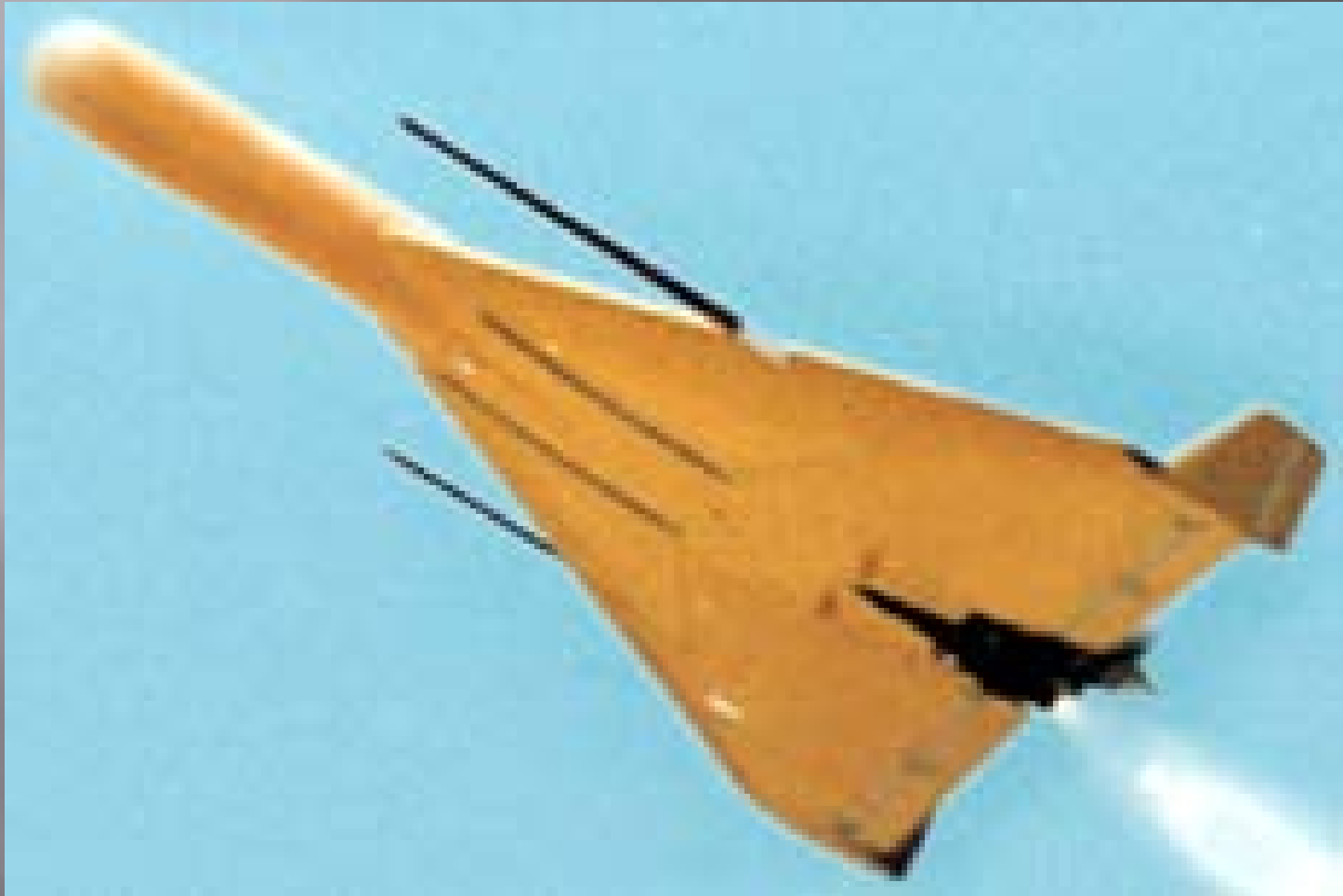
ICRAC - The Scientists' Call

... To *Ban* Autonomous Lethal Robots

As Computer Scientists, Engineers, Artificial Intelligence experts, Roboticists and professionals from related disciplines, we call for a *ban* on the development and deployment of weapon systems in which the decision to apply violent force is made autonomously.

Decisions about the application of violent force *must not* be delegated to machines.

Ban LAR's Poster Child



Potential LAR Scenarios

- ▣ Stupid Algorithm
- ▣ Really Smart Algorithm
 - Comprehensible
 - Black Box (Big Data)
- ▣ Stupid Entity (including savants)
- ▣ Really Smart Entity
 - Benevolent
 - Indifferent to Evil

Algorithm → Entity?

- ▣ Self-reflective
- ▣ Self-modifying (algorithms, not data)
- ▣ Has goals

- ▣ Self-willed?
- ▣ Conscious?
- ▣ *WILL* evolve instrumental subgoals
AKA ethics

Scientists' Call's Reasoning

We are concerned
about the potential of robots
to undermine human responsibility
in decisions to use force, and
to obscure accountability
for the consequences.

Robots or Algorithms?

- ▣ *Death by algorithm is the ultimate indignity*
says 2 star general
- ▣ Ceding godlike powers to robots reduces human beings to things with no more intrinsic value than any object.
- ▣ When robots rule warfare, utterly without empathy or compassion, humans retain less intrinsic worth than a toaster—which at least can be used for spare parts.
- ▣ In civilized societies, even our enemies possess inherent worth and are considered persons, a recognition that forms the basis of the Geneva Conventions and rules of military engagement.

Death by Algorithm – Peter Asaro

While the detailed language defining autonomous weapon systems in an international treaty will necessarily be determined through a process of negotiations, the centrepiece of such a treaty should be the establishment of the principle that

human lives cannot be taken without an informed and considered human decision regarding those lives in each and every use of force,



and any automated system that fails to meet that principle by removing the human from the decision process is therefore prohibited.

A ban on autonomous weapons systems must instead focus on



the delegation of the authority to initiate lethal force to an automated process

not under direct human supervision & discretionary control.

In the near future . . .

- ▣ New York SWAT teams receive “smart rifles”
 - Friendly fire , successful outcomes 
 - “Shoot everything & let the gun sort it out”
 - The rifle is the arbiter of who lives/dies
 - Safety feature turned executioner

In the near future . . .

- ▣ LA SWAT teams introduce “armed telepresence”
 - *Minorly* modified DARPA disaster-relief robots
 - Pre-targeting, aim correction = inhuman speed/accuracy
 - In training exercises, friendly fire , good outcomes 
 - ADD the “smart rifles”?

Summary

Lethal autonomous robotics (LARs) are weapon systems that, once activated, can select and engage targets without further human intervention. They raise far-reaching concerns about the protection of life during war and peace. This includes the question of the extent to which they can be programmed to comply with the requirements of international humanitarian law and the standards protecting life under international human rights law. Beyond this, their deployment may be unacceptable because no adequate system of legal accountability can be devised, and because robots should not have the power of life and death over human beings. The Special Rapporteur recommends that States establish national moratoria on aspects of LARs, and calls for the establishment of a high level panel on LARs to articulate a policy for the international community on the issue.

- ▣ LARs select and engage targets autonomously
 - LARs do *not* include drones (subsequent report)
- ▣ Compliance with IHL and IHRL
- ▣ Adequate system of accountability
- ▣ **Robots** should not have the power of life & death
- ▣ Recommendations – moratoria, policy

UN GA/HRC Report

<among other reasons>

deployment *may* be unacceptable because . . .

robots should not have
the power of life and death
over human beings

Frequently Cited Reasons

- ▣ Fear (Why create something that might exterminate you?)
 - “Terminator” scenario (“Berserkers”)
 - Super-powerful but indifferent
- ▣ Suppresses Democracy
 - What if they end up in the wrong hands?
- ▣ Humanocentrism/Selfishness
 - Right relationship to technology
- ▣ Because it’s a good clear line

Which Is The Problem?

- ▣ Stupid Algorithms
- ▣ Terrifying Entities

- ▣ What if the algorithms were *proven* smarter than 2013 humans?
- ▣ What if the entities were *guaranteed* to be benevolent and altruistic?
(and fully capable of altruistic punishment)

Do we really care about methods or RESULTS?

Engineering for Responsibility

- ▣ Limit any dilution of responsibility
- ▣ Machines must not make 'decisions' that result in the death of humans
- ▣ 'Mala in se' (evil in themselves)
 - Unpredictable/cannot be fully controlled
- ▣ Unpredictability is simply BAD DESIGN
- ▣ Global Hawk UAV had insufficient autonomy
- ▣ More autonomy <> more unpredictability

Intuition Pumps

- ▣ SWAT
 - Lethal but very short-term
 - Lethal but slightly longer term (& more humanoid)
- ▣ Policing
 - Non-lethal but consensus algorithms
 - Non-lethal but big-data algorithms
 - Non-lethal but self-willed
 - Non-lethal, self-willed, wire-headed
- ▣ In evil hands

Wise Strategic Points

- ▣ Never delegate responsibility until recipient is known capable of fulfilling it
- ▣ Don't worry about killer robots exterminating humanity – we will always have equal abilities and they will have less of a “killer instinct”
- ▣ Entities can protect themselves against errors & misuse/hijacking in a way that tools cannot
- ▣ Diversity (differentiation) is *critically* needed
- ▣ Humanocentrism is selfish and unethical

Grounding Responsible Governance in Plausibility/Truth (Scientific “Reality”)

- ▣ Does a proponent truly believe in their position or is it held to pre-empt some slippery slope (a so-called “regulatory Maginot Line”)?
- ▣ Is the proponent honest about their own goals (are they truly transparent)?
- ▣ Do they engage in argumentation and rhetoric or constructive debate?
- ▣ Do they quickly resort to ad hominem and/or accusations of “bad faith” in debate?