

The Ethics & Law of Autonomous Weapons

Peter M. Asaro, PhD

asarop@newschool.edu

Assistant Professor

School of Media Studies, The New School

Affiliate Scholar

Center for Internet and Society, Stanford Law School

Co-Founder & Vice-Chair

International Committee for Robot Arms Control

Overview

- Defining Autonomous Weapon Systems
- Ethical & Legal Problems with AWS
- Meaningful Human Control Solves Multiple Problems
- Software & the Delegation of Authority & Responsibility
 - Policing

Drones & Targeted Killing



175 5.6%

535 17%

2391 75.9%

PAKISTAN

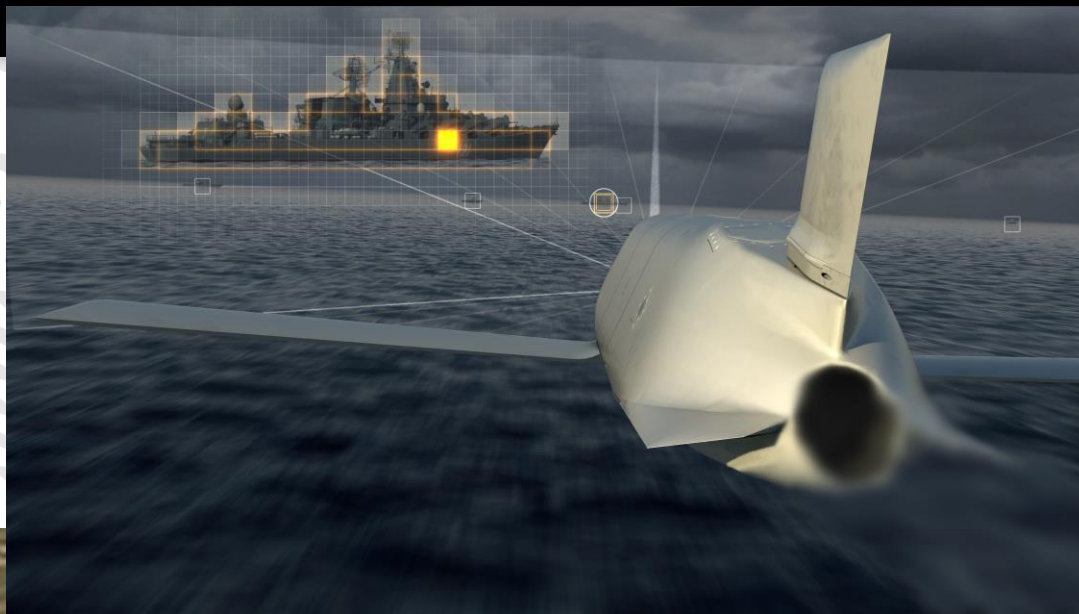
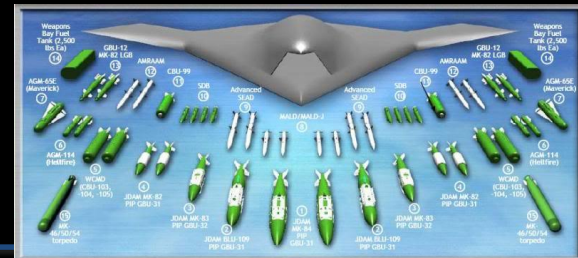
ESTIMATED TOTAL FATALITIES **3149**
SHARE

HIGH PROFILE
1.5% **48**

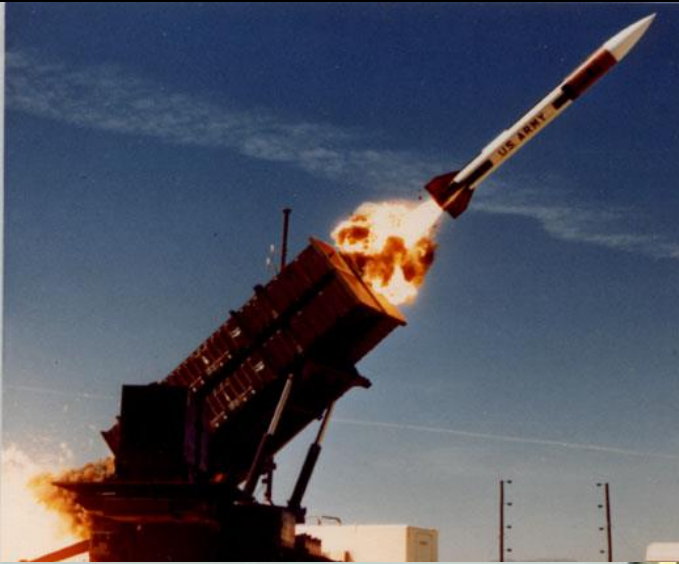
2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013



Military Robotics



Current & Future Weapons



암구호 비정상



Defining Autonomous Weapon Systems

- Lethal Autonomous Weapons Systems (LAWS)
- Fully Autonomous Weapons
- Killer Robots

- Humans-in-the-loop

- Decision Cycle / Kill Chain (USAF)
 - Find
 - Fix
 - Track
 - Target
 - Engage
 - Assess

- Meaningful Human Control

Morality of AWS

- **Consequences**
 - Risks & Consequences
- **Duties**
 - Rights, Responsibilities, Obligations
 - Exercise Moral Agency in Self-Determination
 - Develop Moral Character
- **Virtues**
 - Loyalty
 - Honor
 - Integrity
 - Honesty
 - Courage
- **Sentiments**
 - Principles of Humanity
 - Dictates of Public Conscience

Consequentialist Approaches

- Narrow Consequences
- Wide Consequences
- Compare: Slavery & Torture

Narrow Consequences

- Principle of Distinction
- Proportionality & Respect for Combatants *hors de combat*
- **Negatives:** Fratricide, civilian harm, unintended initiation/escalation of conflict, hacking, spoofing, behavioral hacking, “normal accidents,” common-mode failure, battle-damaged systems, environment
- **Magnitude:** depends on time/space range of weapon, potential damage, size of magazine, time until humans could shut it down, number of like systems deployed, number of “contacts”
- **Failures are a Certainty, Despite Mitigation:**
 - Despite lowering probability of failure, high-frequency use ensures significant number of failures
 - Adversarial Systems are Intrinsically Unpredictable

Autonomy & IHL

- Technical Possibilities
 - Distinction
 - *Hors de combat*
 - Civilians Directly Participating in Hostilities
 - Threshold of Harm
 - Direct Causation
 - Belligerent Nexus
 - Proportionality
 - Military Necessity

Wide Consequences

- **Unintended Initiation & Escalation**
 - Accelerated Pace
 - Decreasing Predictability
- **Arms Races**
 - Regional & Global Instability
 - Waste of Resources
- **Threats to International Humanitarian Law**
 - Unattributable Attacks
 - Accountability
 - Weapons of Mass Destruction

Drivers of Innovation

- What Mechanisms Drive Increased Accuracy/Precision?
- For Whom?
- Proliferation?
- Where does the Data come from?

Human Rights & Dignity

- Is an automated decision to kill morally justifiable?
- Is it ethical to delegate such decision to an AWS?
- How could this undermine international law & human rights?

Justified Killing

- It Matters *How* One is Killed
 - Not Arbitrary, Summary, Extra-Judicial
- Duty of Commanders
 - Take all Reasonable Precautions to Prevent Harms to Civilians & Civilian Infrastructure
 - Assess Military Necessity
 - For Each Attack

Implicit Req. for Human Judgment

Article 57 of Additional Protocol I

1. In the conduct of military operations, **constant care shall be taken** to spare the civilian population, civilians and civilian objects.
2. With respect to attacks, the following precautions shall be taken
 - (a) **those who plan or decide upon an attack shall:**
 - (i) do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects ...
 - (ii) **take all feasible precautions in the choice** of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects;
 - (iii) **refrain from deciding to launch any attack** which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated;

Improper Delegation

- AWS are Not Legal Agents
- AWS are Not Moral Agents
- AWS are Arbitrary/Summary/Extra-Judicial
- Undermines Responsibility
 - Psychology of Human Decisions
 - Legal Accountability

Need for New Law

- Nothing Explicitly Prohibits AWS in Current Law
- Current Weapons Reviews, Wait-and-See, will not prevent unsafe systems from being developed and fielded
- Need a Clear and Strong Norm
- Human Rights & Dignity are firm grounds for New Law

- Sept. 2009: ICRAC Founded
- April 2013: Stop Killer Robots Campaign Launched

- Spring 2014, 2015, 2016: UN CCW Experts Meeting
- Summer 2015, Letter from Scientists, Future of Life Institute
- Spring 2016: CCW Group of Governmental Experts Meeting



Meaningful Human Control

- Positive Control
- Must Allow Human to Exercise Moral Agency
 - Call off an attack when necessary
 - Making Meaning
- Must Provide the Information Necessary for Required Moral Judgments
 - Accuracy
 - Confidence
 - Clarity
- Responsibility: Operators Feel Confidence in Actions Sufficient for Accountability
 - Predictability & Reliability
 - Transparency
- Property of Not Just the Weapon System, but Larger Socio-Technical System
 - Chain of Command
 - Training
 - Maintenance

Policing & Lethal Robots

- Higher Standards for Use of Force
 - “In order to prevent an imminent threat of death or grave bodily harm.”
- Requires:
 - Physical Capability
 - Psychological Intent
- Issues of Racial Bias Remain
- Most Cases are Self-Defense of Officer
- **Answer: No Autonomous Use of Force**

Thank You!

- Email: asarop@newschool.edu
- Twitter: [@peterasaro](https://twitter.com/peterasaro)
- WWW: www.peterasaro.org

Papers

- Asaro, P. (2016). "[Jus nascendi, Robotic Weapons and the Martens Clause](#)," in Ryan Calo, Michael Froomkin and Ian Kerr (eds.) [Robot Law](#), Edward Elgar Publishing, pp. 367–386.
- Asaro, P. (2015) "[Roberto Cordeschi on Cybernetics and Autonomous Weapons: Reflections and Responses](#)," [Paradigmi: Rivista di critica filosofica](#), Anno XXXIII, no. 3, Settembre-Dicembre, 2015, pp. 83-107.
- Asaro, P. (2012). "[On Banning Autonomous Lethal Systems: Human Rights, Automation and the Dehumanizing of Lethal Decision-making](#)," Special Issue on New Technologies and Warfare, [International Review of the Red Cross](#), 94 (886), Summer 2012, pp. 687-709.
- Asaro, P. (2009). "[Modeling the Moral User: Designing Ethical Interfaces for Tele-Operation](#)," [IEEE Technology & Society](#), 28 (1), 20-24.
- Asaro, P. (2008). "[How Just Could a Robot War Be?](#)" in Adam Briggles, Katinka Waelbers and Philip A. E. Brey (eds.), [Current Issues in Computing And Philosophy](#), Amsterdam, The Netherlands: IOS Press, pp. 50-64.

Autonomous Killing

- *Mal en se*
- IHL & HRL require specifically Human decision-making
- The human rights to life and due process, and the limited conditions under which they can be overridden, imply a specific duty with respect to a broad range of automated and autonomous technologies. In particular there is a duty upon individuals and states in peacetime, as well as combatants, military organizations, and states in armed conflict situations, not to delegate to a machine or automated process the authority or capability to initiate the use of lethal force independently of human determinations of its moral and legal legitimacy in each and every case.

Martens Clause & the Limits of the Law

- Art 1(2) of AP I

In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the **principles of humanity** and from the **dictates of public conscience**.

Not Illegal ≠ Moral

Can ≠ Should