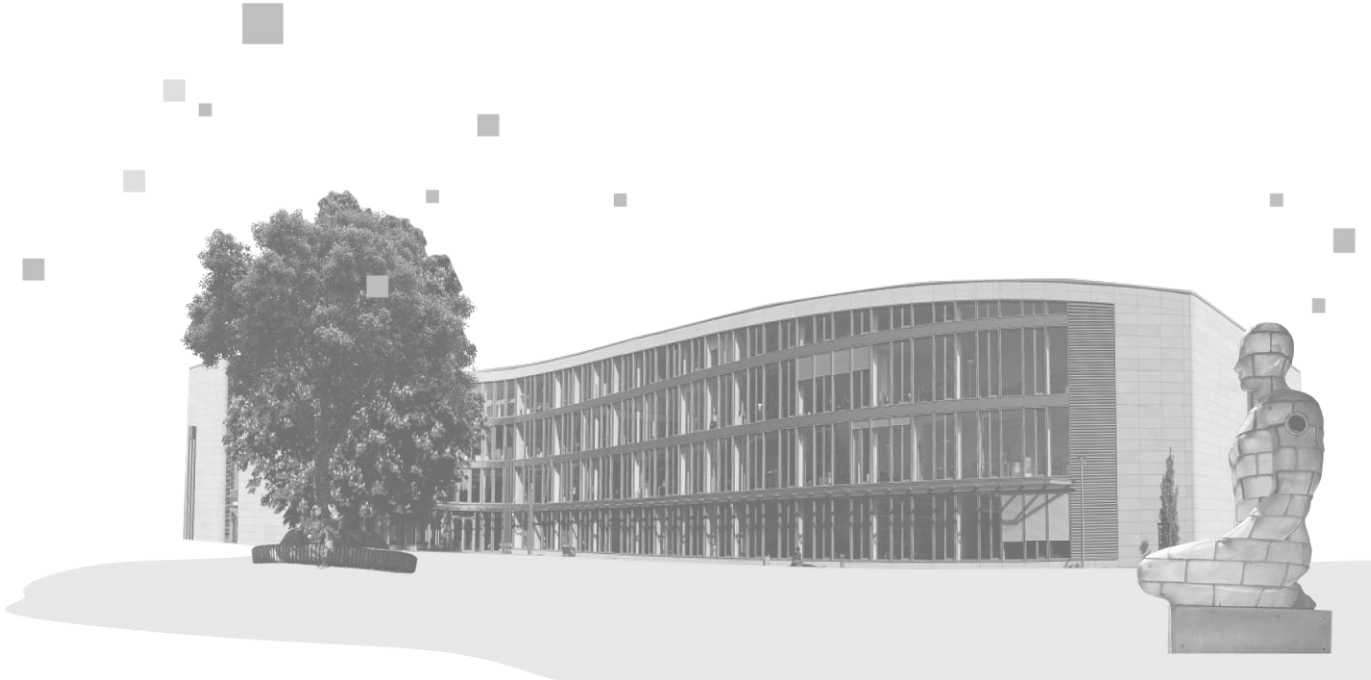
AI in Warfare & Lethal Autonomy

Jörn Sobotta - joern.sobotta@student.hpi.uni-potsdam.de

Accountability in AI



Design IT.
Create Knowledge.

www.hpi.de

Libya 2020

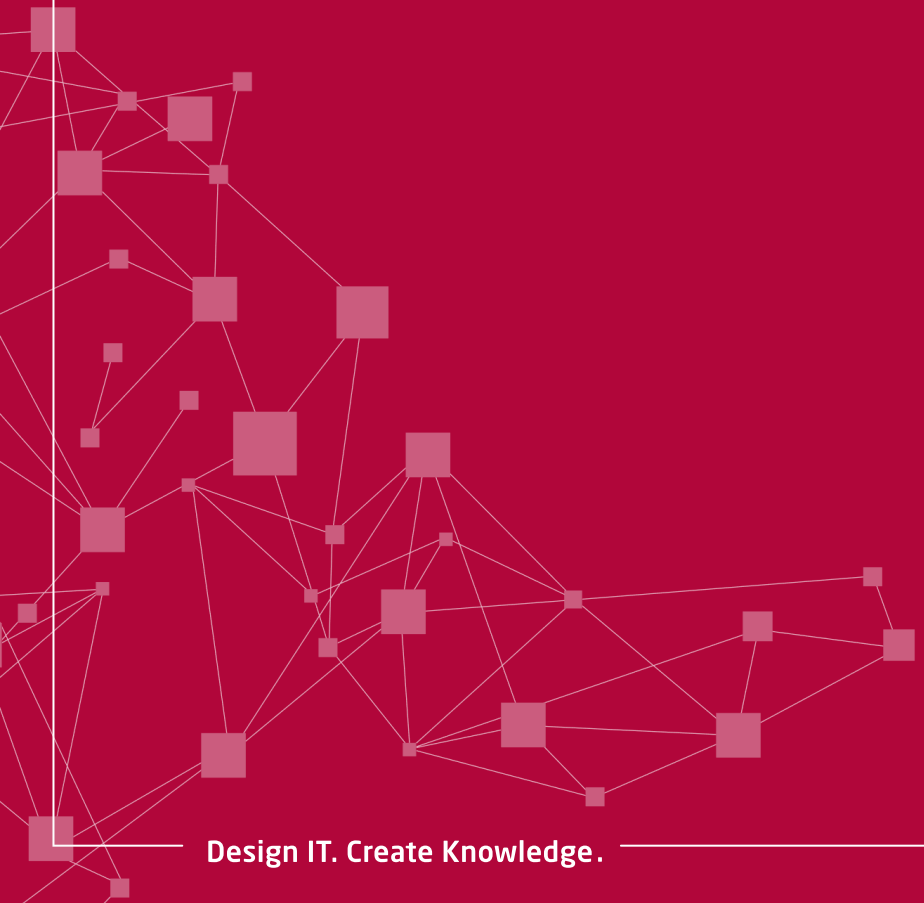
The First ever Kill by Lethal Autonomous Weapon Systems (LAWS)?

near Tripoli:

- UN-backed Government of National Accord (GNA) against Haftar Affiliated Forces (HAF)
- HAF supported by Egypt, Saudi Arabia, UAE, Russia [1]
- GNA supported by Turkey, Qatar, Italy [1]
  - During GNA Counteroffensive "Operation PEACE STORM" with Turkish support, retreating HAF forces hunted down by Kargu-2
    - Programmed to attack targets without requiring data connectivity to operator
    - **"Fire, forget, and find"**
  - **Unclear if people were killed completely autonomously**
  - **Proxy War** (with weak accountability)

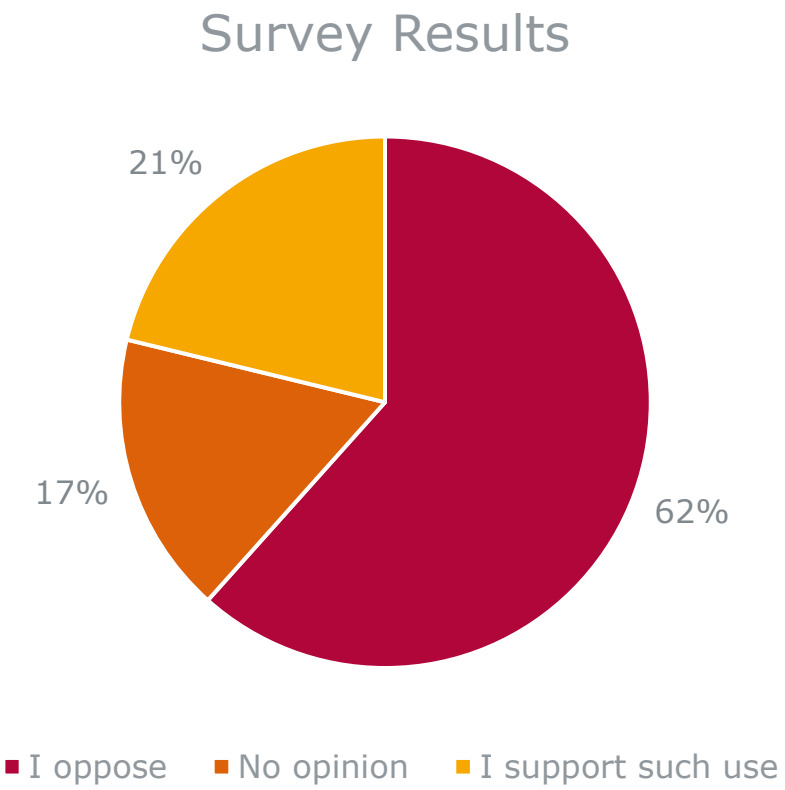
[1] Global Conflict Tracker (2024) (https://www.cfr.org/global-conflict-tracker/conflict/civil-war-libya)
[2] International Committee of the Red Cross (https://casebook.icrc.org/case-study/libya-use-lethal-autonomous-weapon-systems)
[3] United Nations Security Council Report (2021) (https://docs.un.org/en/S/2021/229)

How do you feel about the use of lethal autonomous weapons systems, capable of independently selecting and attacking targets without human intervention, in war?



"[…] How do you feel about the use of such LAWS in war?"[1] [2]

A Study by Ipsos for the Human Rights Watch 2021

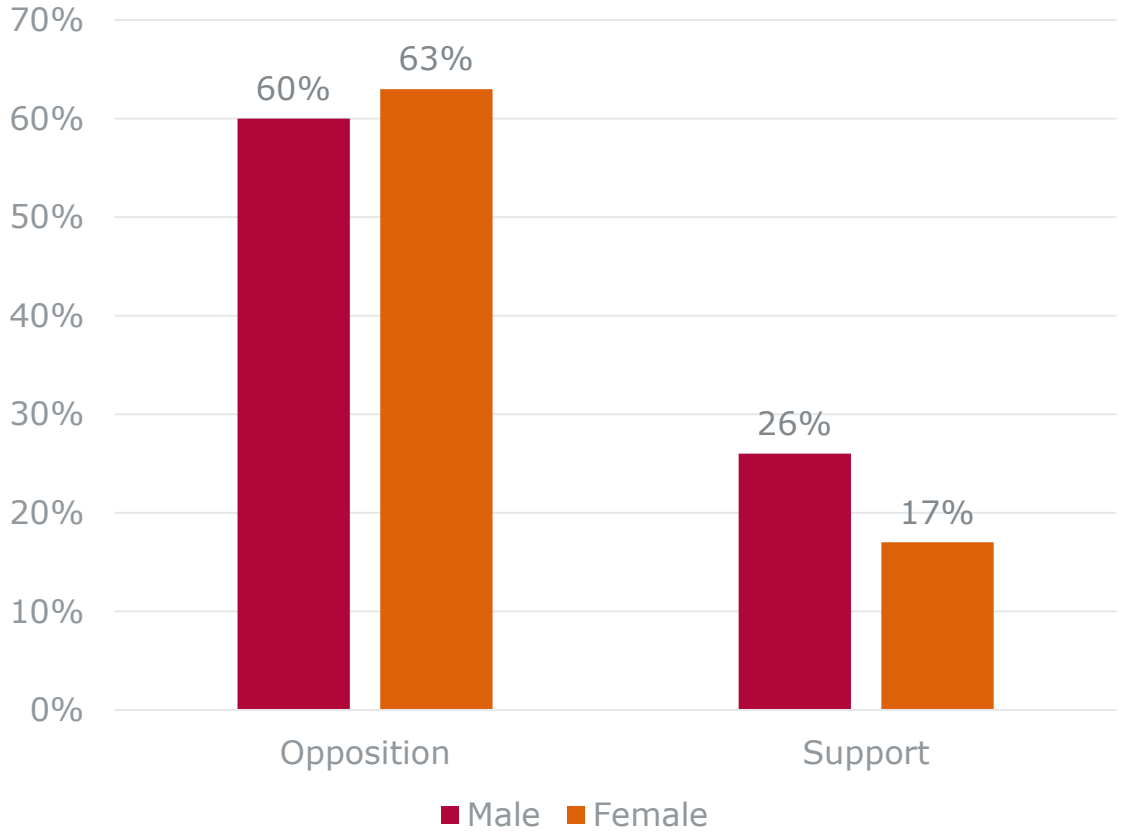


"The United Nations is reviewing the strategic, legal and moral implications of lethal autonomous weapons systems. These weapons systems would be **capable of independently selecting targets** and **attacking those targets without human intervention**. They are thus different than current day "drones" where humans select and attack targets. How do you feel about the use of such lethal autonomous weapons systems in war?"

[1] Human Rights Watch (2021) (https://www.hrw.org/news/2021/02/02/killer-robots-survey-shows-opposition-remains-strong)
[2] Ipsos (2021) (https://www.ipsos.com/en-us/global-survey-highlights-continued-opposition-fully-autonomous-weapons)

"[…] How do you feel about the use of such LAWS in war?"[1] [2]

A Study by Ipsos for the Human Rights Watch 2021



Reasons for opposing:

- 66% say that these systems **cross a moral line**, as machines should not be allowed to kill
- 53% say they are concerned these weapons would be "**unaccountable**"
- 42% worry that killer robots would be subject to **technical failures**
- 24% say they'd be **illegal**

30.04.2026

[1] Human Rights Watch (2021) (https://www.hrw.org/news/2021/02/02/killer-robots-survey-shows-opposition-remains-strong)
[2] Ipsos (2021) (https://www.ipsos.com/en-us/global-survey-highlights-continued-opposition-fully-autonomous-weapons)

(Lethal) Autonomous Weapons Systems - (L)AWS

Dual Use Concern

Human Dignity

Automation Bias

Responsibility Gap

Meaningful Human Control MHC

Lethal Autonomous Weapon Systems

„LAWS are a special class of weapon systems that uses sensor suites and computer algorithms to **independently identify** a target and **employ an onboard weapon system** to engage and destroy the target **without manual human control** of the system.“ [1]

Categories as defined by the Stockholm International Peace Research Institute (SIPRI) [2]:

air defense systems	active protection systems	robotic sentries	guided munitions	loitering munitions
e.g. Iron Dome	protect vehicles e.g. Trophy	stationary robotic platform for surveillance of e.g. demilitarized zone	autonomously identify and engage targets, which are out of range of the attacking aircraft	overfly area in search of targets

[1] U.S. Policy on Lethal Autonomous Weapon Systems (2026) (https://www.congress.gov/crs-product/IF11150)
[2] SIPRI - Mapping the Development of Autonomy in Weapon Systems (2017) (https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf)

(Lethal) Autonomous Weapons Systems - (L)AWS

Dual Use Concern

Human Dignity

Automation Bias

Responsibility Gap

Meaningful Human Control MHC

WHO – What is dual-use research of concern? (2020) (https://www.who.int/news-room/questions-and-answers/item/what-is-dual-use-research-of-concern)

(Lethal) Autonomous Weapons Systems - (L)AWS

Dual Use Concern

Human Dignity

Automation Bias

Responsibility Gap

Meaningful Human Control MHC

Human Rights Watch – A Hazard to Human Rights (2025) (https://www.hrw.org/report/2025/04/28/a-hazard-to-human-rights/autonomous-weapons-systems-and-digital-decision-making)

(Lethal) Autonomous Weapons Systems - (L)AWS

Dual Use Concern

Human Dignity

Automation Bias

Responsibility Gap

Meaningful Human Control MHC

Bundeszentrale für politische Bildung (2025) (https://www.bpb.de/lernen/digitale-bildung/werkstatt/569392/automation-bias/)
EY – AI Sentiment Index 2025 (https://www.ey.com/de_de/newsroom/2025/05/ey-ai-sentiment-index-2025)

(Lethal) Autonomous Weapons Systems - (L)AWS

Dual Use Concern

Human Dignity

Automation Bias

Responsibility Gap

Meaningful Human Control MHC

(Lethal) Autonomous Weapons Systems - (L)AWS

Dual Use Concern

Human Dignity

Automation Bias

Responsibility Gap

Meaningful Human Control MHC

The Concept of MHC – Meaningful Human Control [1]

defined by Article 36

1. A machine applying force and operating without any human control whatsoever is broadly considered unacceptable.
2. 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.

[1] Article 36 - Meaningful Human Control, Artificial Intelligence and Autonomous Weapons (2016) (https://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf)

The AWS Ethical and Legal Debate

Ronald C. Arkin vs. Noel Sharkey

Arkin: Ethical Pros of Autonomy

"In the fog of war it is hard enough for a human to be able to effectively discriminate whether or not a target is legitimate. […] [F]uture autonomous robots may be able to perform better than humans" [1]

- **No self-preservation as foremost drive**
- **"Scenario fullfillment" [3]**
- **Can be designed without emotions**

Sharkey: Ethical Cons of Autonomy

"[Challenges are] establishing responsibility for war crimes involving autonomous weaponry, the potential lowering of the threshold for entry in war, the military's reluctance to give robots the right to refuse an order, […], cybersecurity, [] and mission creep." [2]

[1] Arkin RC. Governing Lethal Behavior in Autonomous Robots (2009)(https://www.cs.cmu.edu/~illah/CLASSDOCS/Arkin.pdf)
[2] Arkin RC. Lethal Autonomous Systems and the Plight of the Non-combatant (2014) (https://d-nb.info/1052151809/34)
[3] EBSCO - USS Vincennes Shotts Down Iranian Civilian Plane (2023) (https://www.ebsco.com/research-starters/military-history-and-science/uss-vincennes-shoots-down-iranian-civilian-plane)

The four ethical and legal concerns

by Christof Heynes – UN Human rights Council (2013) [1]

Compliance with IHL

Responsibility Ascription Problems

Violations of Human Dignity

Increased Risk for Peace and International Stability

IHL: International Humanitarian Law

[1] Christof Heynes, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions (2013) (https://digitallibrary.un.org/record/755741)

The AWS Ethical and Legal Debate

International Regulations

United Nations Convention on Certain Conventional Weapons (CCW) [1]

Restrict and possibly ban the use of weapons which may be deemed to be excessively injurious or affect civilians indiscriminately.

Group of Governmental Experts on emerging technologies in the area of Lethal Autonomous Weapons Systems (GGE on LAWS) [2]

Non-binding 11 Guiding Principles on LAWS

Lack of Common Purpose:

"Russia [...] opposes any new rules or regulations" [3]

[1]The Convention on Certain Conventional Weapons (https://disarmament.unoda.org/en/our-work/conventional-arms/convention-certain-conventional-weapons)
[2] CCW: GGE on LAWS - EU LTTs (2024) (https://www.eeas.europa.eu/delegations/un-geneva/ccw-gge-laws-eu-ltts_en)
[3] Human Rights Watch - Agenda Action (2022) (https://www.hrw.org/report/2022/11/10/agenda-action/alternative-processes-negotiating-killer-robots-treaty)

The AWS Ethical and Legal Debate

Further Opposition and Guidelines

Future of Life Institute: Open Letter on autonomous Weapons [1]

(4.500 AI/robotics researchers + 26.000 others):

"chemists and biologists have broadly supported international agreements that have successfully prohibited chemical and biological weapons, just as most physicists supported the treaties banning space-based nuclear weapons and blinding laser weapons."

Human Rights Watch: Stopping Killer Robots (2020) [2]

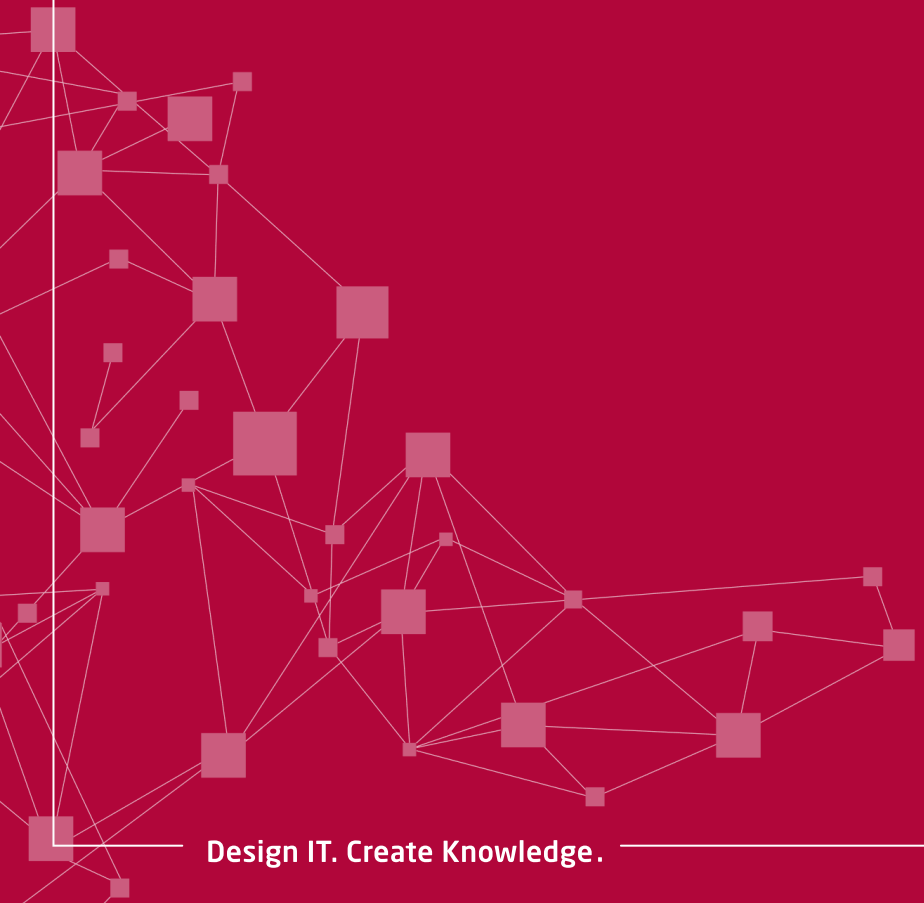
30 countries have called for a ban on fully autonomous weapons (e.g. Austria, China (only deployment, not production) and Iraq)

Guideline Proposal: Human Rights Watch: A Hazard to Human Rights (2025) [3]

[1] Future of Life Institute – Autonomous Weapons Open Letter (2016) (https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/)
[2] Human Rights Watch - Stopping Killer Robots (2020) (https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and)
[3] Human Rights Watch – A Hazard to Human Rights (2025) (https://www.hrw.org/report/2025/04/28/a-hazard-to-human-rights/autonomous-weapons-systems-and-digital-decision-making)

Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues

Daniele Amoroso, Guglielmo Tamburrini



Policy Proposals

Uniform Policies for Meaningful Human Control

One policy to fit all AWS

Boxed Autonomy Policy

- **Constraining the autonomy of weapons systems within an operational box**
- **Predifined parameters, fixed time period and geographical borders**
- **Boundaries set by human operator**

Deliberate Targeting

Denied Autonomy Policy

- **Most restrictive interpretation of MHC**
- **Criticized for setting too high a threshold for machine autonomy**
- **Too far off of reality of warfare**

Unfeasible for e.g. air defense system

Supervised Autonomy Policy

- **Requires humans to be in the loop of AWS**
- **Provides human operators with the ability to intervene and terminate engagements**

But with faster offensive AWS human operators in the loop only becomes an illusion

Automation Bias

Differentiated Policies for Meaningful Human Control

Proposed by Noel Sharkey[1], refined by Amoroso & Tamburrini

Level 1	Level 2	Level 3	Level 4	Level 5
A human engages with and selects targets and initiates any attack	A program suggests alternative targets, and a human chooses which to attack	A Program selects targets, and a human must approve before the attack	A Program selects and engages targets but is supervised by a human who keeps the power to override its choices and abort the attack	A program selects targets and initiates attacks on the basis of the mission goals as defined at the planning/activation stage, without further human involvement
No Autonomy	Support System		Supervised Autonomy	
		Automation Bias		Boxed Autonomy

[1] International Committee for Robot Arms Control (2018) (https://www.icrac.net/wp-content/uploads/2018/04/Sharkey_Guideline-for-the-human-control-of-weapons-systems_ICRAC-WP3_GGE-April-2018.pdf)

Differentiated Policies for Meaningful Human Control

Proposed by Noel Sharkey[1], refined by Amoroso & Tamburrini

Level 1	Level 2	Level 3	Level 4	Level 5
A human engages with and selects targets and initiates any attack	A program suggests alternative targets, and a human chooses which to attack	A Program selects targets, and a human must approve before the attack	A Program selects and engages targets but is supervised by a human who keeps the power to override its choices and abort the attack	A program selects targets and initiates attacks on the basis of the mission goals as defined at the planning/activation stage, without further human involvement
ideal	acceptable	unacceptable		

[1] International Committee for Robot Arms Control (2018) (https://www.icrac.net/wp-content/uploads/2018/04/Sharkey_Guideline-for-the-human-control-of-weapons-systems_ICRAC-WP3_GGE-April-2018.pdf)

My own Opinion

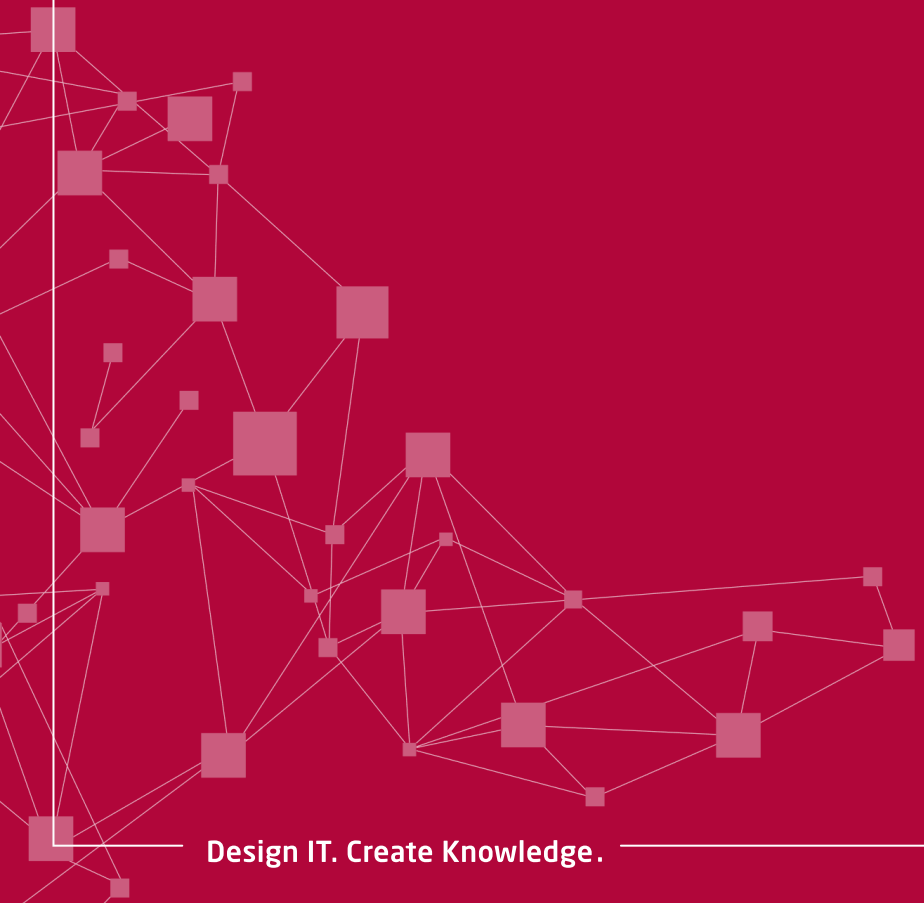
Compliance with IHL

Responsibility Ascription Problems

Violations of Human Dignity

Increased Risk for Peace and International Stability

How do you feel about the use of lethal autonomous weapons systems, capable of independently selecting and attacking targets without human intervention, in war?



- If it could be proven that AWS make fewer targeting errors than stressed human soldiers, would that change your position?

- Is there a meaningful difference between being killed by a soldier who hates you, a soldier who feels nothing, and an algorithm that classified you as a target?

- The countries blocking a treaty are also the ones most advanced in developing AWS. If international regulation is structurally impossible, what realistic alternatives exist? Or do we simply accept an unregulated arms race?