ZAGREB APPLIED ETHICS CONFERENCE 2017
THE ETHICS OF ROBOTICS AND ARTIFICIAL INTELLIGENCE
Matica hrvatska - ZAGREB, CROATIA
5-7 JUNE 2017

Session II – 14:00

The viability of 'embedded Ethics' in robotic military systems without humans in the decision loop

Miguel Moreno Muñoz
University of Granada (Spain)
mm3@ugr.es

Aim and objectives

Aim: To analyze the viability of “embedded Ethics”, as an alternative to “machine Ethics” or “engineering safety” approaches concerning to Artificial Intelligent agents.

Objectives

- ✓ Review the literature about machine Ethics and engineering safety
- ✓ Develop a case-study research on autonomous weapons
- ✓ Explore the prospect for an “embedded Ethics” approach
- ✓ Use the military context as a benchmark
- ✓ Extrapolate results and conclusion to civil/industrial AI agents

THE FUTURE OF WARFARE

The future of warfare

- A. Krishnan (2014): http://robohub.org/robots-soldiers-and-cyborgs-the-future-of-warfare

- **Public perception associated with anthropomorphic-cyborg soldiers**
  - The future of warfare: a battlefield where humanoid robots and other machines fight alongside or in the place of human soldiers.
    - Droids of Star Wars
    - The Terminator's cyborg soldiers
  - In the real world, robots are simply programmable machines that can sense and interact with their environment.
    - Most advanced weapons systems are robotic, including cruise missiles, drones, and air and missile defense systems.
    - Are we at the beginning of an inevitable process leading to the rise of "killer robots"?, or can robots actually make war less destructive?

Limitations for exoskeleton or *land warriors*

- A. Krishnan (2014): http://robohub.org/robots-soldiers-and-cyborgs-the-future-of-warfare

- **The *Cyborgization* of Human Soldiers has clear limits:**
  - Robotic weapons systems could be combined with humans, whose bodies could be augmented with robotic technology.
    - This concept offers the quick reaction times, precision, and strength of robotic systems, and the control and superior cognitive abilities of humans.
  - DARPA's Land Warrior and its successor projects (Objective Force Warrior, Future Force Warrior, Warrior Web)
    - aiming to equip soldiers with wearable computers, advanced communications gear, helmet visors with night vision and head-up-display, and robotic exoskeletons to improve mobility.
    - Such human enhancement had some setbacks and slow progress (as other robotic systems): The gear is still too heavy, and the exoskeletons that could enable soldiers to carry more and move faster lack a sufficient power source.

      → There are many reasons to exclude humans from battlefields.

Advances in autonomous helicopters

http://www.foxnews.com/us/2014/04/06/new-technology-to-enable-navy-drones-to-choose-flight-paths-landing-sites.html

- **Use of drones in U.S. military operations**
  - To counter the improvised explosive device (IED) threat in Iraq
  - To carry out aerial bombing campaigns in Afghanistan, Pakistan, Yemen, Somalia
- **The "truly leap-ahead technology" is to get advances in autonomous helicopters**
  - "Full size helicopters able to deliver 5,000 pounds of cargo"
    - Helicopters that can choose their own routes, pick landing sites and change their destinations if they spot unexpected obstacles that emerge at the last minute.
    - Reduce the need to use ground convoys to deliver food, water, and weapons.
      - Ground convoys are attractive targets for enemy fighters.
      - From 2003 to 2007, one person was killed or wounded for every 24 fuel resupply convoys in Afghanistan and one was killed or wounded for every 29 water resupply convoys.

Fast growing demand

Sehrawat, V. (2017). Autonomous weapon system: Law of armed conflict (LOAC) and other legal challenges. *Computer Law & Security Review*, *33*(1), 38–56. http://doi.org/10.1016/j.clsr.2016.11.001

- There are over 12,000 robots on the ground and 7,000 in the air in these conflict areas.
  - In 2003, the U.S. had no ground robots in Iraq or Afghanistan;
- In Iraq, robots have defused over 10,000 roadside bombs, responsible for 40% of U.S. casualties there.
  - In the next decade, the military is aiming to create autonomous aircraft that can help soldiers carry out night raids, search oceans or forests and select targets for attack.
  - “Human beings want their gadgets to cook, clean, read, dictate, count, and solve problems for them. Now, humans must decide if they want gadgets to fight for them as well” (p. 39).

Human control over “kill missions” under threat

https://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html

- **Budget cuts means less human oversight**
  - For the moment, the Pentagon's expanding drone fleet has limited ability to operate autonomously.
    - Autonomous drones that require less human oversight could also take some strain off the Pentagon as it cuts back the size of the military to deal with budget cuts.
      - The Navy's drone the X-47B, landed itself on an aircraft carrier (July 10, 2013), is still experimental.
    - The Army plans to create a robot that can operate on its own in helping soldiers search for suspects.
      - New drones have limited capabilities (flying at night, in difficult weather).
      - Those obstacles should be overcome by 2016-2018.

Joint Strike Missile are, in fact, “killer robots”?

https://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html

- **Some weapons are explicitly designed to permit human operators to step away from controls**
  - Israel’s antiradar missile (Harpy) loiters in the sky until an enemy radar is turned on. It then attacks and destroys the radar installation on its own.
  - Norway plans to equip its fleet of advanced jet fighters with the *Joint Strike Missile*, which can hunt, recognize and detect a target without human intervention.
    - Weapons that make their own decisions move so quickly that human overseers soon may not be able to keep up.
    - Smarter weapons should be embraced because they may result in fewer mass killings and civilian casualties. “They do not commit war crimes”. (Paul Scharre, NYT)

The first generation of drones is obsolete

http://robohub.org/robots-soldiers-and-cyborgs-the-future-of-warfare

- **IED hunters and unmanned aerial vehicles**
  - Northrop Grumman's Global Hawk surveillance drones
  - General Atomics' armed Predator and Reaper drones
- **Inability to survive in contested airspaces** and general budgetary pressures
  - Reorientation to a Prompt Global Strike (PGS) program (to develop the capability to attack any target within one hour worldwide).
  - At the moment, the only PGS weapons currently available to the U.S. are severely restricted nuclear missiles.
- The foreseeable trend: development of robotic systems for shorter but more massive air campaigns against more sophisticated adversaries with modern air defense systems, like Libya, Syria, China, or Russia.

Tracking responsibility when AWS are used

Schulzke, M. (2013). Autonomous weapons and distributed responsibility. *Philosophy and Technology*, *26*(2), 203–219. http://doi.org/10.1007/s13347-012-0089-0

- **AWS deployed on the battlefields could obfuscate who should be held responsible for how these weapons act**
  - **Robert Sparrow:** it would be impossible to attribute responsibility for autonomous robots' actions to their creators, their commanders, or the robots themselves.
  - **Marcus Schulzke:** the problem of determining responsibility for autonomous robots can be solved by addressing it within the context of the military chain of command.
    - The military hierarchy is a system of distributing responsibility between decision makers on different levels and constraining autonomy.
    - If autonomous weapons are employed as agents operating within this system, then responsibility for their actions can be attributed to their creators and their civilian and military superiors.

Cooperative missiles aren't science fiction

John Markoff, Fearing Bombs that Can Pick Whom to Kill

https://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html

Credit: Defense Advanced Research Projects Agency

– **Long Range Anti-Ship Missile prototype**, launched by a B-1 bomber, is **designed to maneuver without human control**.

  - Without human oversight, the missile decided which of three ships to attack, dropping to just above the sea surface and striking a 260-foot unmanned freighter.
  - Warfare is increasingly guided by software: weapons that rely on artificial intelligence, not human instruction, to decide what to target and whom to kill, are being developed.

Cooperative missiles aren't science fiction

John Markoff, Fearing Bombs that Can Pick Whom to Kill

https://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html

– **Autonomous weapons rely on artificial intelligence and sensors to select targets and to initiate an attack.**
  - Britain, Israel and Norway are already deploying missiles and drones that carry out attacks against enemy radar, tanks or ships without direct human control.

– Britain's "fire and forget" Brimstone missiles can distinguish among tanks and cars and buses without human assistance
  - It can hunt targets in a predesignated region without oversight.
    – The Brimstones also communicate with one another, sharing their targets.
    – Armaments with even more advanced self-governance (to avoid fire) are on the drawing board.

– "An autonomous weapons arms race is already taking place,"
  - Steve Omohundro: "They can respond faster, more efficiently and less predictably."

⟶ Wang, F. B., & Dong, C. H. (2013). Fast Intercept Trajectory Optimization for Multi-stage Air Defense Missile Using Hybrid Algorithm. *Procedia Engineering*, *67*, 447–456.

Cooperative missiles aren’t science fiction

John Markoff, Fearing Bombs that Can Pick Whom to Kill

https://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html

– **Technological advances in three particular areas have made self-governing weapons a real possibility**.

- New types of radar, laser and infrared sensors are helping missiles and drones better calculate their positions and orientations.
- “Machine vision,” resembling that of humans, identifies patterns in images and helps weapons distinguish important targets.

- Representatives from dozens of nations met in Geneva to consider whether development of these weapons should be restricted by the Convention on Certain Conventional Weapons (13-17 April 2015).
- Christof Heyns, the United Nations special rapporteur on extrajudicial, summary or arbitrary executions, called for a moratorium on the development of these weapons.
- The Pentagon has issued a directive requiring high-level authorization for the development of weapons capable of killing without human oversight. But fast-moving technology has already made the directive obsolete

FINAL STAGE TOWARDS LETHAL AUTONOMOUS WEAPONS SYSTEMS

International Committee for Robot Arms Control

http://icrac.net/2016/11/new-icrac-leaflet-on-laws-and-global-security

- **A group of scientists that advocates restrictions on the use of *Lethal Autonomous Weapons Systems* (LAWS)**
  - Autonomous Weapons Systems (LAWS): weapons that once activated will select targets and attack them with violent force without the benefits of human control.
    - Are these human-designated targets?
    - Are these systems automatically deciding what is a target?
    - ICRAC is pushing for an international legally binding treaty to prohibit the development production and use of LAWS.
      - ICRAC is worried about the destabilizing impact that LAWS will have on the security of the planet.
      - 'LAWS take the automation of weapon systems a step too far, undermining the conditions necessary for meaningful human control.'

Race is in the final stage towards LAWS

Shawn Helton, Hollywood Comes Real: The Future of Warfare Will be 'Decided by Drones' Not Humans, 21st Century Wire (Oct. 8, 2013)

- Research in "AI will soon allow drones to perform targeted killing without the consultation of their human masters, working autonomously, coldly responding to a set of predetermined criteria." (Shawn Helton, Cent. Wire)
- "Scientists, engineers and policymakers are all figuring out ways drones can be used better and more smartly, more precise and less damaging to civilians, with longer range and better staying power. One method under development is by increasing autonomy on the drone itself." (Joshua Foust, Nat. J.).

LAWS – 10 Problems for Global Security

http://icrac.net/wp-content/uploads/2015/04/LAWS-10-Problems-for-Global-Security.pdf

- **Specific risks of *Autonomous Weapons Systems* (LAWS)**
  - *The end point of increasing weapons' automation is full autonomy, where human beings have little control over the course of conflicts and events in battle. At this point in time, it is still within our power to stop the automation of the kill decision, by ensuring that every weapon remains meaningfully controlled by humans.*
  - *Both humans and computer systems have their strengths and weaknesses, and the aim of designing effective supervisory systems for weapons control must be to exploit the strengths of both. This way, it is possible not only to gain better legal compliance, but also to ensure that the partnership between human and machine best ensures the protection of civilians, their human dignity and our wider global security.* (p. 2)

Humans can’t stand the pace of present battles

ICRAC (2015). *LAWS: Ten Problems For Global Security*. Geneva (p. 2, § 5).
http://icrac.net/wp-content/uploads/2015/04/LAWS-10-Problems-for-Global-Security.pdf

- New prototypes in the unmanned systems domain are increasingly being tested at supersonic and hypersonic speeds.
  - This will require even faster autonomous response devices that in turn will require ever-faster weapons.
  - Such a ‘pace race’ will equate to humans having little control over the battle-space.
- “**Meaningful human control**” is something under discussion:
  - “If a drone’s system is sophisticated enough, it could be less emotional, more selective and able to provide force in a way that achieves a tactical objective with the least harm.” (Samuel Liles)
  - “A lethal autonomous robot can aim better, target better, select better, and in general be a better asset with the linked ISR [intelligence, surveillance, and reconnaissance] packages it can run.” (S. L.)

    → Drones have built-in vulnerabilities to viruses and hijacking with inexpensive equipment (http://edition.cnn.com/2012/07/19/us/house-drones-hacking-risk/)

Could a robot be a *moral reasoner*?

Pontier, M. A., & Hoorn, J. F. (2012). Toward machines that behave ethically better than humans do. In *Proceedings of of the 34th International Annual Conference of the Cognitive Science Society* (pp. 2198–2203).

– **Pontier & Hoorn: Some machines can behave ethically better than humans do.**

- viability of a moral reasoner that combines connectionism, utilitarianism and ethical theory about moral duties?
  - The moral decision-making matches the analysis of expert ethicists in the health domain, where machines interact with humans (clinical context).
  - But: Moral decision making is arguably even one of the most challenging tasks for computational approaches to higher-order cognition.
- **Pros**: the behavior of machines is still far easier to predict than the behavior of humans.
  - human behavior is typically far from being morally ideal
  - humans are not very good at making impartial decisions
  - machines capable of sufficient moral reasoning would even behave ethically better than most human beings would.

Is *Machine Ethics* a Wrong Approach?

Allen, C., Wallach, W., & Smit, I. (2006). Why Machine Ethics? *IEEE Intelligent Systems*, *21*(4), 12–17. http://doi.org/10.1109/MIS.2006.83

– **Allen, Wallach & Smit are optimist:**

- Humans have always adapted to our technological products, and the benefits of having autonomous machines will most likely outweigh the costs. But optimism doesn't come for free.
- We already have semiautonomous robots and software agents that violate ethical standards as a matter of course:
  - A search engine might collect data that's legally considered to be private.
  - Public concerns regarding the future takeover of humanity by a superior form of AI or the havoc created by endlessly reproducing nanobots.
  - Only in crisis situations beyond the scope of any programming, human judgment would be preferred, and would involve ethical considerations.

- ***Artificial Moral Agents*** (AMAs) should be able to make decisions that honor privacy, uphold shared ethical standards, protect civil rights and individual liberty, and further the welfare of others.

  → "Good" is defined and measured against specific purposes of designers and users. Artificial morality: ways of getting artificial agents to act as if they were moral agents (Allen at al., 13).

Ethical decisions framed as safety problems

Allen, C., Wallach, W., & Smit, I. (2006). Why Machine Ethics? *IEEE Intelligent Systems*, *21*(4), 12–17. http://doi.org/10.1109/MIS.2006.83

- **Allen et al. suggest to defer questions about whether a machine can be genuinely ethical**
  - Even to be genuinely autonomous presume that a genuine ethical agent acts intentionally, autonomously, and freely.
    - But the present engineering challenge concerns only artificial morality:

      → ways of getting artificial agents to act as if they were moral agents.
    - What is necessary to trust multipurpose machines, programmed to respond flexibly in real or virtual environments?

      → This means **something more than traditional product safety**?
  - "If an autonomous system is to minimize harm, it must be cognizant of possible harmful consequences and select its actions accordingly." (p. 13)
    - It doesn't exclude the "engineering safety approach".

Is *Machine Ethics* a Wrong Approach?

Yampolskiy, R. V. (2013). Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach (pp. 389–396). http://doi.org/10.1007/978-3-642-31674-6_29

- **Yampolskiy: the attempts to allow machines to make ethical decisions are misguided:**
  - Proposes a new science of safety engineering for intelligent artificial agents. The goal: to prove that they are in fact safe even under recursive self improvement.
  - The driverless trains today are ethically oblivious, but safe.
    → Should software engineers attempt to enhance their software systems to explicitly represent ethical dimensions of situations?
    → Could it be applicable to military contexts?
  - I suggest to consider "embedded Ethics" as an alternative:
    - Instructions added to the program or control software coherent with agreed results after a careful analysis of foreseeable options, taken into account specialized ethical criteria.
    - In extreme cases, those instructions could prevent wrong or illegal orders from human agents, at the risk of subverting the chain of command.

THE ENGINEERING-SAFETY APPROACH

Artificial Intelligence Safety Engineering

Yampolskiy, R. V. (2013). Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach (pp. 389–396). http://doi.org/10.1007/978-3-642-31674-6_29

- **Yampolskiy:** human-like moral performance means some immoral actions, not acceptable from the machines we design.
  - **Robin Hanson**: "In the long run, what matters most is that we all share a mutually acceptable law to keep the peace among us, and allow mutually advantageous relations, not that we agree on the "right" values. Tolerate a wide range of values from capable law-abiding robots. It is a good law we should most strive to create and preserve. Law really matters." (Prefer Law to Values, Oct. 10, 2009)
- **From artificial moral agents to *capable law-abiding robots***
  - **Yampolskiy** proposes that purely philosophical discussions of ethics for machines be supplemented by scientific work aimed at creating safe machines in the context of a new field he will term "AI Safety Engineering."
  - Unfortunately, **International Humanitarian Law (IHL) or Law of Armed Conflict (LOAC) does not currently have provisions for LAWS** (it is unclear whether the international community would be supportive of a treaty that would limit or ban lethal autonomous robots).
  - The **Convention on Conventional Weapons**, which is grounded in IHL, provides an important framework to further understanding of the technical, legal, ethical and policy questions raised by the development and use of LAWS in armed conflicts.

Legal compliance in war is not only Safety Engineering

Sehrawat, V. (2017). Autonomous weapon system: Law of armed conflict (LOAC) and other legal challenges. *Computer Law & Security Review*, *33*(1), 38–56. http://doi.org/10.1016/j.clsr.2016.11.001

- **LAWS are safe in some critical aspects, but not *confined***
  - Without a legal framework robust enough to prevent violations of International Humanitarian Law (IHL), Law of Armed Conflict (LOAC) or the Convention on Conventional Weapons, there is a risk of escalation in the race for developing the most lethal system.
- **“Safety engineering” is a relational term, when used in military context:**
  - the intended safety is only for the users/owners controlling the deployment of the lethal system, not for their potential targets.
- **Conclusion**: The **safety engineering approach is not adequate** for socio-technical scenarios where *Lethal Autonomous Weapon Systems* (LAWS) can be deployed.

Moral responsibility for attack, not only for malfunction

http://www.unesco.org/new/en/social-and-human-sciences/themes/comest/science-ethics/robotics-questions-and-answers/
http://www.unesco.org/new/en/media-services/single-view/news/unesco_science_experts_explore_robots_rights/

- **Industrial robotics is developed on a clear delimitation of responsibilities, but still remains under regulated.**
  - The complexity of contemporary robots arises questions about who should bear responsibility for malfunction/harm to human beings.
    - **UNESCO: Ethically and legally, robotics remains under regulated**
      - » It is a relatively new and rapid changing field of research whose impact on the real world is often difficult to anticipate.
      - » There are no specific ethical guidelines as to how robotic research and projects, with direct impact on humans, should proceed.
      - » There are no universally accepted codes of conduct for robotics. However, robots are treated in the same way as any other technological product in terms of legal regulation.
    - **Autonomous robotic cars or armed military robots** could go out of control. “The question is, therefore, not only if roboticists ought to respect certain ethical norms, but whether certain ethical norms need to be programmed into the robots themselves.”

Is Artificial General Intelligence (AGI) research unethical?

Yampolskiy, R. V. (2013). Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach (pp. 389–396). http://doi.org/10.1007/978-3-642-31674-6_29

- **Responsibility is sometimes difficult to be traced**
  - For AI safety engineering, the grand challenge is to develop **safety mechanisms for self-improving systems**.
    - An artificially intelligent machine, as capable as a human engineer of designing the next generation of intelligent systems, should have a safety mechanism incorporated in the initial design, and still functional after thousands of generations of continuous self-improvement without human interference.
    - Ideally, every generation of self-improving system should be able to produce a verifiable proof of its safety for external examination.
  - **R.V. Yampolskiy**: It would be catastrophic to allow a safe intelligent machine to design an inherently unsafe upgrade for itself resulting in a more capable and more dangerous system.
    - **Artificial General Intelligence** (AGI) research should be considered unethical.

Malfunction caused by bugs or random events

Yampolskiy, R. V. (2013). Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach (pp. 389–396). http://doi.org/10.1007/978-3-642-31674-6_29

- **Yampolskiy: AI safety engineering faces major challenges in complex systems.**
  - **Ted Kazynsky**: Machine-made decisions will bring better result than man-made ones. Bugs, random or unexpected events can be catastrophic.
    - Eventually a stage may be reached at which the decisions necessary to keep the system running will be so complex that human beings will be incapable of making them intelligently. At that stage the machines will be in effective control.
    - People won't be able to just turn the machines off, because they will be so dependent on them that turning them off would amount to suicide.
    - In general a machine should never be in a position to terminate human life or to make any other non-trivial ethical or moral judgment concerning people.
    - A world run by machines will lead to unpredictable consequences for human culture, lifestyle and overall probability of survival for the humankind.

VIABILITY OF "EMBEDDED ETHICS"

Ethical design for war and peace

Armin Krishnan, http://robohub.org/robots-soldiers-and-cyborgs-the-future-of-warfare

- **Advanced artificial intelligence could develop in directions not anticipated by scientists**.
  - Unpredictability justifies to **preserve humans in the decision loop**.
    - Unmanned weapons systems will become gradually more autonomous to carry out very specific missions with less human direction.
    - They may never entirely replace human soldiers on the battlefield.
    - Unclear role that robots could play in non-traditional wars such as the war on terror or the fight against transnational organized crime.
  - Great doubts about the military usefulness and legality of Autonomous military robots, except when a fast sequence of actions is needed.
    - Adverse budgetary climate will focus the research and acquisition on the projects relevant to the most likely military needs, but won't stop the race.

      → designing robots for war may be a secondary priority

      → cyborgizing soldiers and further developing human-operated robotic systems couldn't be discarded, but it is not the most likely scenario.

The prospect for an embedded Ethics

Allen et al. (2006) vs. Yampolskiy (2013)

- **Intelligent Artificial Agents could be safe without to explicitly represent ethical dimensions of situations**
  - In military context, the pace of battle often demands self-governing machines, capable of assessing the ethical acceptability of the options they face.
    - Clear distinction between combatants and non-combatants is often impossible.
- Only recently designers consider the ways in which they implicitly embedded values in the technologies they produced.
    - Ethicists can help engineers to become aware of their work's ethical dimensions.
    - More attention to unintended consequences resulting from the adoption of information technology.
    - Attention to the values that are unconsciously built into technology.
  - But, the morality implicit in artificial agents' actions is simply a question of engineering ethics? (→ engineers recognizing their ethical assumptions)
    - Modern devices' complexity is associated with increased difficulty in predicting the outcome.
    - The modular design of systems can mean that no single person or group can fully grasp the manner in which the system will interact or respond to new inputs.

Care ethics for service robots – some elements

van Wynsberghe, A. 2016. "Service Robots, Care Ethics, and Design." *Ethics and Information Technology* 18 (4): 311–21. 10.1007/s10676-016-9409-x

- **Amee van Wynsberghe: Care Centered Value Sensitive Design (CCVSD)**
  - A robust proactive framework for incorporating ethics into the design and implementation of robots (health care use).
    - There will be approx 35 million service robots at work by 2018.
    - Their design and development demand ethical attention.
    - Normative foundations for CCVSD: the care ethics tradition and in particular the use of care practices for: (1) structuring the analysis and, (2) determining the values of ethical import.
    - Some examples of current robot prototypes that can and cannot be evaluated using CCVSD

Rational Behavior Model (RBM) architecture

Brutzman, D., McGhee, R., & Davis, D. (2012). An Implemented Universal Mission Controller with Run Time Ethics Checking for Autonomous Unmanned Vehicles-a UUV Example. In 2012 IEEE/OES AUTONOMOUS UNDERWATER VEHICLES (AUV). 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE.

- **Brutzman et al**. developed and tested **a three-level software architecture called Rational Behavior Model** (RBM), in which a top (strategic) level mission control finite state machine (FSM) orders the rational execution, at an intermediate (tactical) level, of vehicle behaviors, to carry out a specified mission.
  - Based on this experience, they believe that human-like intelligence and judgment are not required to achieve a useful operational capability in autonomous mobile robots.
- According to them, a primitive but useful type of robot ethical behavior can also be attained, even in hazardous or military environments, without invoking concepts of artificial intelligence.
  - **Key feature**: mission orders can be tested exhaustively in human executable form before being translated into robot executable form.
  - This provides the kind of transparency and accountability needed for after action review of missions, and possible legal proceedings in case of loss of life or property resulting from errors in mission orders.

Unpredictability and human reasoning

Heyns, C. (2013). *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*. http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf

- **Lethal Autonomous Weapons Systems (LAWS) could be a threat to life protection during war and peace.**
  - Once activated, LAWS can select and engage targets without further human intervention.
    - It is under discussion 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.
    - Without an adequate system of legal accountability, the UN Special Rapporteur recommends that States establish national moratoria on aspects of LAWS, and the establishment of a high level panel to articulate a policy for the international community on the issue.
    - Robots should not have the power of life and death over human beings.

→ OHCHR: Evaluate the legal, ethical and policy issues related to LARs (and drones), to ensure transparency, accountability and the rule of law.

Mobility, interactivity, communication and autonomy

UNESCO. (2016). *Preliminary Draft Report of COMEST on Robotics Ethics*. Paris. http://unesdoc.unesco.org/images/0024/002455/245532E.pdf

- **Certain features typically associated with contemporary robots raise unique ethical concerns**
  - crucial for understanding what robots are, individually or jointly:
    → mobility, interactivity, communication and autonomy.
  - **Robots can be stationary (most industrial robots are) or mobile**
    - **mobility** is essential for many types of robots because it allows them to perform tasks in place of humans in typically human environments (e.g. hospital, office or kitchen).
      - Robot mobility can be realized in various technical ways: able to walk (bipedal and multilegged robots), crawl, roll, wheel, fly and swim.
    - The range of possible harm caused by stationary robot is limited to those working or living in its proximity
    - Mobile robots (especially if they have advanced autonomy and capacity to interact with their environment) may pose more serious threats.

Autonomy in “critical functions”

ICRC: https://www.icrc.org/en/document/lethal-autonomous-weapons-systems-LAWS

An autonomous weapon system (LAWS) is a lethal system that has autonomy in its ‘critical functions’:

- **Can select** (i.e. search for or detect, identify, track) **and attack** (i.e. intercept, use force against, neutralize, damage or destroy) **targets without human intervention**
  - It is important to know how autonomy is developed in these ‘critical functions’
  - Particularly in ‘targeting decision-making’, and therefore to compliance with international humanitarian law and its rules on distinction, proportionality and precautions in attack.
  - Autonomy in the critical functions of selecting and attacking targets raise significant ethical questions when force is used autonomously against human targets.

Autonomy in "critical functions"

ICRC: https://www.icrc.org/en/document/lethal-autonomous-weapons-systems-LAWS

The **ICRC suggests to ground discussions on current and emerging weapon systems** that are pushing the boundaries of human control over the critical functions.

- Hypothetical scenarios about possible developments far off in the future could neglect autonomy in the critical functions of weapon systems that actually exist or will be deployed in the near future.
- **Many of the existing autonomous weapon systems have autonomous 'modes'**, and therefore only operate autonomously for short periods. They also tend to be highly constrained in the tasks they are used for, the types of targets they attack, and the circumstances in which they are used.
- Most existing systems are also overseen in real-time by a human operator.

Degrees of artifactual morality?

Dodig Crnkovic, G., & Çürüklü, B. (2012). Robots: ethical by design. *Ethics and Information Technology*, *14*(1), 61–71. http://doi.org/10.1007/s10676-011-9278-2

- **Highly intelligent and autonomous artifacts with significant impact and complex modes of agency must be equipped with more advanced ethical capabilities.**
  - Artificial morality is considered to be the ability of a machine to perform activities that would require morality in humans.
  - The capacity for artificial (artifactual) morality, such as artifactual agency, artifactual responsibility, artificial intentions, artificial (synthetic) emotions, etc., comes in varying degrees and depend on the type of agent.
- **Artificial moral agents have functional responsibilities** within a network of distributed responsibilities **in a socio-technological system**
  - **This does not take away the responsibilities of the other stakeholders** in the system, but facilitates an understanding and regulation of such networks.
  - The process of development must assume **an evolutionary form with a number of iterations because the emergent properties of artifacts** must be tested in real world situations.

Conclusion:
Humans (still) should be in the decision loop

- **Standard attribution of responsibility in industrial robotic, assist. / health (controlled environments)**
  - Engineers, manufacturers, programmers, qualified operators…
  - Safety and embedded ethics approaches

- **For military operations in open environments**
  - Unpredictability could harm civilians and noncombatants
  - Emergent properties of AI-robotic systems deployed
  - Specific challenges to use LAWS under international humanitarian law criteria, in complex scenarios
  - So far, AI is applied knowledge to specific purposes, in a sequence of actions initiated by humans (the only morally capable agents).