Autonomous Weapons and Moral Distance

Laura Nolan

António Guterres
@antonioguterres

Autonomous machines with the power and discretion to select targets and take lives without human involvement are politically unacceptable, morally repugnant and should be prohibited by international law. bit.ly/2JGExMD

804 6:28 PM - Mar 25, 2019

458 people are talking about this

UNDER SECRETARY OF DEFENSE FOR INTELLIGENCE
officium nostrum est adiuvare
ALGORITHMIC WARFARE CROSS-FUNCTIONAL TEAM

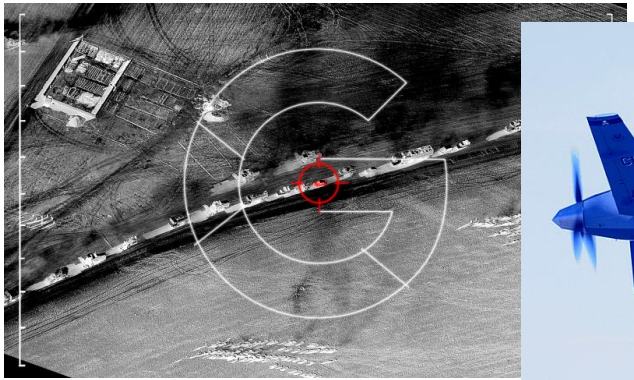
Establishment of Project Maven

- April 2017: Robert O. Work, deputy US secretary of defense, establishes the ‘Algorithmic Warfare Cross-Functional Team’ (a.k.a. Maven) in a memo
- Its stated mission: “accelerate DoD’s integration of big data and machine learning” by working with industry
- Its first project: provide computer vision algorithms for turning drone video footage into ‘actionable intelligence and insights’
- Then moving into ‘other defense mission areas’

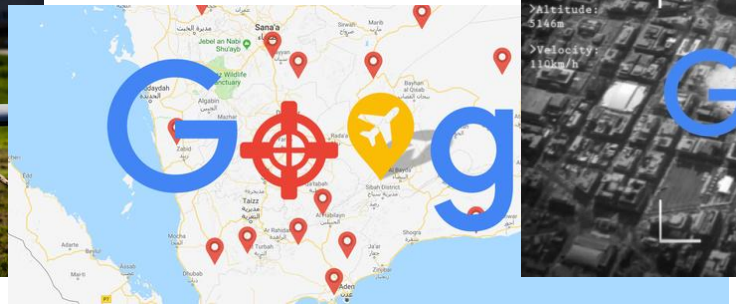
“This program is truly about increasing the lethality of our department.”

--DoD Chief Management Officer John Gibson

DON'T BE EVIL
End the work on the Pentagon's drone project



Google


Sana'a
Taizz
Aden

FLIR
>Altitude:
5146m
>Velocity:
110km/h
Google

- **Sensor To Shooter**: This will build on Maven to develop algorithms that can shrink the time to locate potential targets, prioritize them, and present them to a human, who will decide what action to take. In keeping with Pentagon policy, Shanahan assured me, "this is about making humans faster, more efficient, and more effective. Humans are still going to have to make the big decisions about weapons employment."

Why are people's decisions sometimes worse with computer support?

Eugenio Alberdi[1], Lorenzo Strigini[1], Andrey A. Povyakalo[1], and Peter Ayton[2],

[1]Centre for Software Reliability, City University London, London, UK
[2]Psychology Department, City University London, London, UK
{e.alberdi, strigini, povyakalo}@csr.city.ac.uk, P.Ayton@city.ac.uk

"Automation is increasingly taking on the role of supporting knowledge-intensive human tasks rather than directly replacing some of the human's functions. This actually makes the problem of computer-related human errors subtler."

Maven and the Military Kill Chain

- Target identification: Automated via Maven and Sensor to Shooter
- Force dispatch to target
- Decision to attack: Semi-automated with no strategy articulated to mitigate automation bias
- Destruction of target

So why did *I* oppose Maven?

Diego Delso, delso.photo,
License CC-BY-SA

“No machines were sold – only leased. IBM was the sole source of all punch cards and spare parts. It serviced the machines on site either directly or through its authorized dealer network or field trainees. There were no universal punch cards. Each series of cards was custom-designed by IBM engineers to capture information going in and to tabulate information the Nazis wanted to extract.”

—Edwin Black

“We have long worked with government agencies to provide technology solutions. This specific project is a pilot with the Department of Defense, to provide open source TensorFlow APIs that can assist in object recognition. … The technology flags images for human review, and is for non-offensive uses only.”

-- Google spokesperson

“The experience of the Holocaust beings into relief, however, another social mechanism; one with a much more sinister potential of involving in the genocide a much wider number of people who never in the process face consciously either difficult moral choices or the need to stifle inner resistance of conscience. The struggle over moral issues never takes place, as the moral aspects of actions are not immediately obvious or are deliberately prevented from discovery and discussion.”

Zygmunt Bauman, *Modernity and the Holocaust*

The Intercept_

GOOGLE HIRED GIG ECONOMY WORKERS TO IMPROVE ARTIFICIAL INTELLIGENCE IN CONTROVERSIAL DRONE-TARGETING PROJECT

Lee Fang

February 4 2019, 6:32 p.m.

DONATE →

22

Michael Gruijters at Dutch Wikipedia
License CC-BY-SA

“Inhibitions against violent atrocities tend to be eroded once three conditions are met, single or together; the violence is authorized […]; actions are routinized (by rule-governed practices and exact specification of roles); and the victims of violence are dehumanized.”

Zygmunt Bauman, *Modernity and the Holocaust*

“People and computers will work symbiotically to increase the ability of weapon systems to detect **objects**.”

-- Drew Cukor, chief of the DoD’s Algorithmic Warfare Cross-Function Team

Surgeonsmate at English Wikipedia
License CC-BY-SA 3.0

[War ethos] ‘seems to be a matter of distance and technology. You could never go wrong if you killed people at long range with sophisticated weapons’.

-- Philip Caputo

They performed their calculations and spoke their strange and esoteric tongues because to do otherwise would be to recognize, all too clearly and constantly, the ghastliness of their contemplations. They contrived their options because without them the bomb would appear too starkly as the thing that they had tried to prevent it from being but that ultimately it would become if it ever were used—a device of sheer mayhem, a weapon whose cataclysmic powers no one really had the faintest idea of how to control.

Fred Kaplan, *The Wizards of Armageddon*

[C]laims of precision targeting, and counterfactual narratives about casualties in hypothetical alternative scenarios play an important role in American public discourses about military intervention, with new technologies often presented as magically salvationist actors in the drama. The preeminent example is the conventional American defense of the atomic bombing of Hiroshima and Nagasaki in which the use of the atomic bomb is said to have ended a brutal war and avoided the bloodletting of an otherwise inevitable US land invasion of Japan.

-- Hugh Gusterson, ‘Drone Warfare in Waziristan and the New Military Humanism’

There has been public focus on a limited contract we entered into in September 2017 that fell under the U.S. Department of Defense's Maven initiative. This contract involved drone video footage and low-res object identification using AI, saving lives was the overarching intent.

Diane Greene, CEO of Google Cloud

It'll save lives, they said.

The problem [of accuracy in weapons] of course, is that this encourages commanders to increase their target sets to include objects that are deeply embedded in civilian areas, greatly increasing the risk to noncombatants.

-- Bruce Cronin, *Bugsplat*

Risk transfer [from belligerent combatants to civilians] is an important component of the Western way of war inasmuch as it enables states to initiate military action without fear of incurring politically damaging casualties among their soldiers.

-- Bruce Cronin, *Bugsplat*

US drone warfare in Waziristan has been legitimated through a discourse of military humanism that claims very low rates of civilian casualties and a concern to spare the lives of the innocent. In practice, in concert with the Pakistani government's counterinsurgency campaign and the tactics of the Taliban, drone strikes in Waziristan have killed substantial numbers of civilians and, in a manner reminiscent of the effects of death squads in Central and Latin America, have torn apart Waziri civil society while creating a culture of terror. "Drone essentialism" (a false conviction that drones are inevitably used in a way that minimizes suffering) has concealed a process of "ethical slippage" through which drone operators relaxed their operational practices.

-- Hugh Gusterson, 'Drone Warfare in Waziristan and the New Military Humanism'

ПОБЕДА!

KARGU

Autonomous Rotary Wing
Attack Drone

Technical problems

- Autonomous weapons are likely to be unpredictable
  - These will be very complex systems operating in very complex environments
  - Many autonomous weapons are likely to incorporate machine learning algorithms that are ‘black boxes’
  - Comprehensive testing is impossible - warfare is not a ‘closed world’ problem and it is not stable
- Attacks must comply with the Laws of Armed Conflict - but proportionality requires judgment and understanding of context
- Ethics is not finite and definable as a set of technical guidelines
- Automation bias - meaningless human ‘control’

Daniel Lawrence Lu Wikipedia
License CC-BY-SA 4.0

CAMPAIGN TO STOP
KILLER ROBOTS
StopKillerRobots.org

Greenland
Iceland
Finland
Sweden
Norway
Russia
United Kingdom
Germany
Ukraine
Kazakhstan
Mongolia
France
Spain
Italy
Turkey
China
South Korea
Japan
Canada
United States
Mexico
North Pacific Ocean
North Atlantic Ocean
Iraq
Iran
Afghanistan
Pakistan
Algeria
Libya
Egypt
Saudi Arabia
India
Thailand
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
Kenya
DRC
Tanzania
Angola
Namibia
Botswana
Madagascar
South Africa
Indonesia
Papua New Guinea
Australia
New Zealand
Indian Ocean
South Atlantic Ocean
South Pacific Ocean
Venezuela
Colombia
Peru
Brazil
Bolivia
Chile
Argentina

The UK's position

UK definition of autonomous weapons systems, UK Ministry of Defence (MOD) Joint doctrine note (2011):

> "An autonomous system is capable of understanding higher level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be."

Elements of a Treaty

- A general obligation to maintain meaningful human control over the use of force
- Prohibitions on weapons systems that select and engage targets and by their nature pose fundamental moral or legal problems
    - In particular, systems that rely on data to represent humans
- Specific positive obligations to help ensure that meaningful human control is maintained in the use of all other systems that select and engage targets

Get campaign updates | Donate | Contact your government | Spread the word | Collaborate | Upcoming events

https://www.stopkillerrobots.org/act/