

# AI & Robotics

Ethical aspects

# Goals 1/2



## The junior-colleague

- can describe 10 ethical aspects of AI in own words.
- can explain the link between the technology/humanity balance and the failure or success for all of us.
- can explain why (current) AI/technology has no ethics.
- can describe the term ethics in own words.
- can explain the link between AI and GDPR.
- can think critically about the ethical aspects of killerbots.
- can describe in simple terms the belgian killerbots law.
- can explain the implications of hacking on devices/robots.
- can describe the (possible) impact of robots/AI on the workforce, scientific community, ...
- can explain the implications of face news & the (possible) link with AI.
- can provide some repercussions already happened in the world.
- can describe the role of open source in AI and the ethical “limits”.
- can explain the 9 ethical issues of AI according to the WEF.
- can explain the 3 levels of moral agency.
- can describe the difference between roboethics and machine ethics.

# Goals 2/2



## The **junior-colleague**

- can write down Asimov's three laws of Robotics.
- can think critically about the ethical aspects of Human-Robot interaction
- can think critically about the ethical implications of Artificial Intelligence
- can think critically about the future of technology

# Why?

**Privacy**

**Impact on the workforce  
(Replacement?)**

Generation gap

Psychological problems

**(Geo)politics**

**Hacking**

Digital divide

**Fake news**

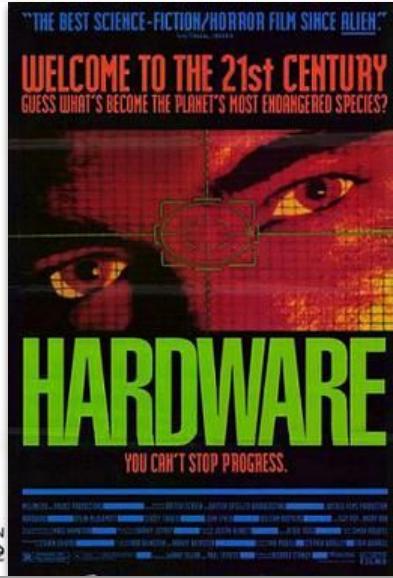
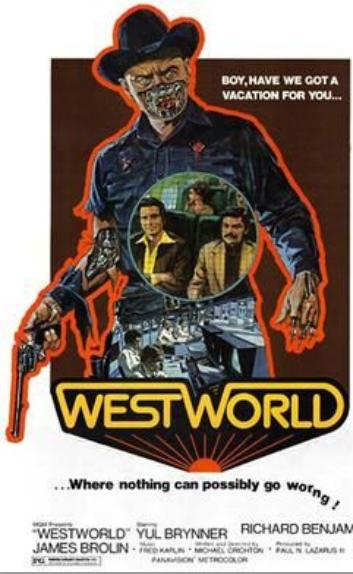
**Killerbots**

**Accidents  
(Who's to blame?)**

**(Lack of?) legislation**

**Impact on the scientific community**

# Why?



# Why?

*The primitive forms of artificial intelligence we already have, have proved very useful. But I think the development of full artificial intelligence could spell the end of the human race.*

-- Professor Stephen Hawking, BBC 2014

# Why?

*AI is a rare case where we need to be proactive about regulation instead of reactive. Because I think by the time we are reactive in AI regulation, it's too late.*

*AI is a fundamental risk to the existence of human civilization in a way that car accidents, airplane crashes, faulty drugs or bad food were not – they were harmful to a set of individuals within society, of course, but they were not harmful to society as a whole*

-- Elon Musk at the National Governors Association 2017

# Why?

*I am in the camp that is concerned about super intelligence. First the machines will do a lot of jobs for us and not be super intelligent. That should be positive if we manage it well. A few decades after that though the intelligence is strong enough to be a concern.*

*I agree with Elon Musk and some others on this and don't understand why some people are not concerned.*

*-- Bill Gates on Reddit 2015*

# The future: utopian or dystopian?



**Finding the right balance between technology and humanity  
will determine failure or success for all of us**



[source] Futurist / Humanist Gerd Leonhard.

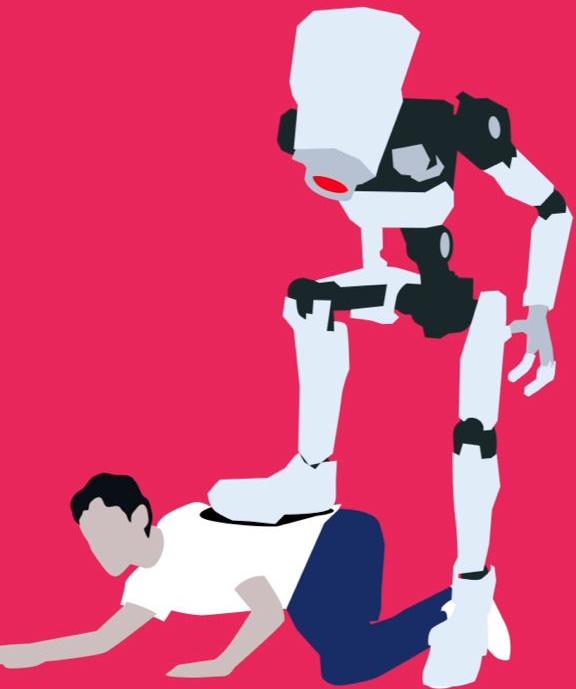
Gerd



[source] Futurist / Humanist Gerd Leonhard.

Gerd

# The future is better than we tend to think....



## TEAM ROBOT



## TEAM HUMAN

[source] Futurist / Humanist Gerd Leonhard.

Gerd

Soon, the question will no longer be if or how technology can do something, but why we should do it, and who



**TECHNOLOGY HAS NO ETHICS**

[source] Futurist / Humanist Gerd Leonhard.

Gerd

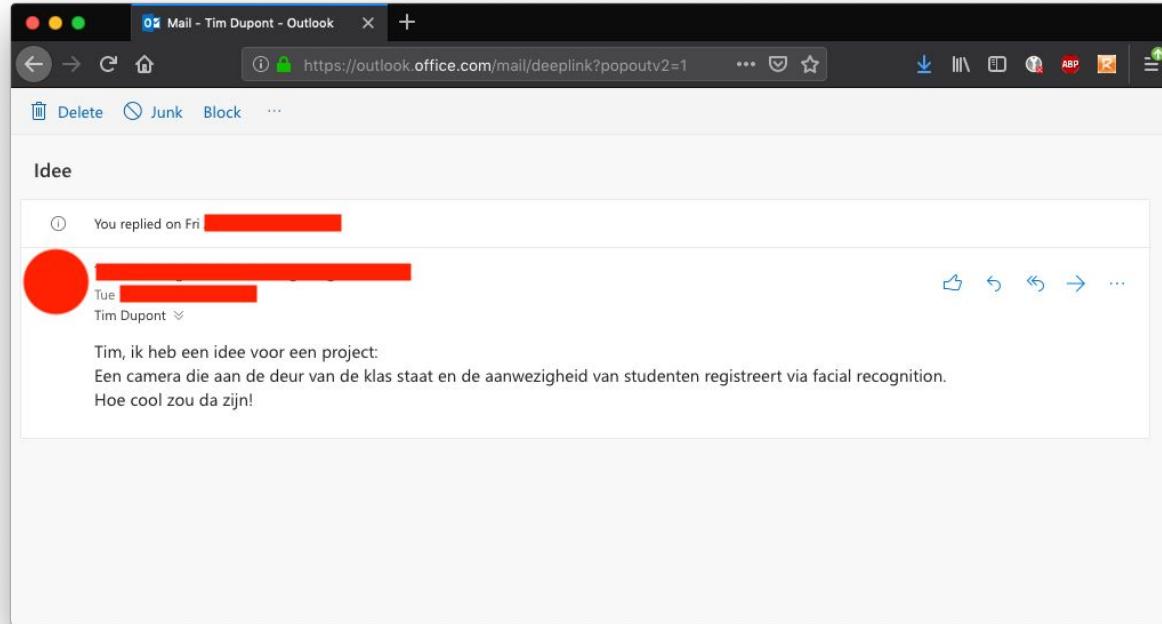




**"Ethics is knowing the difference between what you have a right (or the power) to do and what is the right thing to do"**

adapted from Potter Stewart

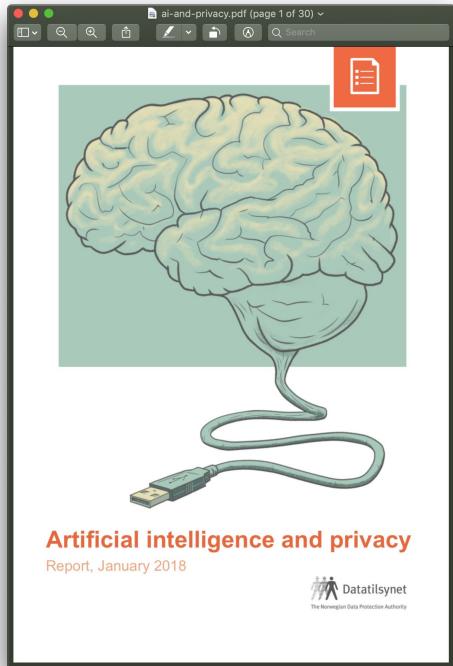
# Project idea from a colleague



# Privacy



# Norwegian Data Protection Authority rapport '18



## Recommendations for developers of AI

Conduct research into how intelligent systems can be made more privacy friendly, such as how AI systems can be designed in order to make it easy for users to comply with the regulations. Research can, for example, be carried out on solutions that use less training data, anonymisation techniques and on solutions that explain how systems process data and how they reach their conclusions. Other interesting research areas include how to conduct system audits to ensure the system isn't biased, especially audits by third parties

Adopt a multidisciplinary approach. AI is more than just technology. It is important to put together multi-disciplinary teams that can consider the consequences for society of the systems developed. Research can also throw light on the how the use of AI can be of considerable value to society as well as on the problematical areas.



The 'TRITON' – under development by Northrop Grumman for the US Navy to be part of the Broad Area Maritime Surveillance (BAMS) programme – is an advanced intelligence, surveillance and reconnaissance missions system, which may operate under control or **autonomously**.  
(Courtesy of Northrop Grumman)

**FICTION**

Wide Field Cameras

Face Recognition

Tactical Sensors

Shaped Explosive





[source] <https://milremrobotics.com>



AUTONOMOUS SECURITY

Self-Driving Technology | Robotics + Artificial Intelligence

Isn't It Time You Gave Your Security Program a 21st-Century Overhaul?

U.S. Deployments Available Nationwide

# Rejected internship

**Afstudeerrichting:** Applicatie-ontwikkeling

**Opdracht**

[REDACTED] Autonomous Hitman

**Omschrijving**  
België heeft onlangs het gebruik van "killer robots" verboden, waarschijnlijk met het idee dat zo'n technologie nog niet voor morgen is. Wij willen graag bewijzen dat de toekomst al een stuk dichterbij staat door onze [REDACTED] om te vormen tot een

---

met waterpistool gewapende huurmoordenaar. Door gebruik te maken van vrij beschikbare open source technologie zoals facial recognition, SLAM en depth estimation willen we [REDACTED] toelaten om autonoom een ruimte te navigeren, de target te identificeren en vervolgens te raken met een waterpistool.

**Doelstelling**  
Het onderzoek beslaat volgende onderdelen:

- Onderzoek naar de Simultaneous Localisation And Movement mogelijkheden van [REDACTED] om een ruimte te navigeren
- Onderzoek naar light weight facial recognition algoritmes, die bij voorkeur uit één enkele foto een doelwit kan identificeren.
- Gebruik maken van Depth Estimation of de twee camera's [REDACTED] om betrouwbaar een doelwit op afstand te raken
- Indien nodig een video streaming architectuur opzetten om [REDACTED] te ontlasten van zware AI berekeningen.

**Voorkennis:**

# Belgium was the first to say NO!

NIEUWS > NIEUWS > ARCHIEF

## België eerste land dat killerrobots verbiedt

05 juli 2018 00:00

□ f in tw m

De Kamercommissie-Landsverdediging heeft een resolutie goedgekeurd die de regering vraagt erover te waken dat het Belgisch leger tijdens militaire operaties geen gebruikmaakt van volledig autonome wapens, zogenaamde killerrobots. De tekst vraagt ook dat de regering ijvert voor een internationaal verbod op die wapens. België is het eerste land dat zo'n verbod instelt. Volledig autonome wapens bestaan nog niet. De robots kunnen zonder menselijke tussenkomst doelwitten dodelijk aanvallen.

# International campaign

The screenshot shows the homepage of the Campaign to Stop Killer Robots website. At the top left is the logo "CAMPAIGN TO STOP KILLER ROBOTS" with a gear icon. The top right features a "Select Language" dropdown, a "DONATE" button, and social media links for Instagram, Facebook, Twitter, and YouTube. The main banner has a dark background with blurred lights and the text "WHO WANTS TO BAN FULLY AUTONOMOUS WEAPONS?". Below the banner, there's a graphic of a flag with the number "5" and the word "Bolivia" underneath it. At the bottom, there are three cards: "The Problem & The Solution", "Global poll shows 61% oppose Killer Robots", and "Rise of the tech workers".

CAMPAIGN TO STOP  
KILLER ROBOTS

LEARN ACT ABOUT NEWS

A A Select Language ▾

DONATE

WHO WANTS TO BAN FULLY AUTONOMOUS WEAPONS?

5

Bolivia

The Problem & The Solution

Global poll shows 61% oppose Killer Robots

Rise of the tech workers

[source] <https://www.stopkillerrobots.org>

# International campaign

TPF

ABOUT WORK NEWS SERVICES TEAM CONTACT

X



HBO DOCUMENTARY FILMS

HBO

THE TRUTH ABOUT KILLER ROBOTS

*A work of science non-fiction*

Documentary, 2018

Writing in 1942 and foreseeing a time when robots are everywhere, science fiction writer Isaac Asimov devised a series of laws to protect humans from robots. Asimov's first law was: "A robot may not injure a human being or, through inaction, allow a human being to come to harm." The future that Asimov envisioned is upon us.

[source] <http://www.thirdpartyfilms.com>

# 1979: Robert Williams (First human killed by a robot.)



[Info] [https://en.wikipedia.org/wiki/Robert\\_Williams\\_\(robot\\_fatality\)](https://en.wikipedia.org/wiki/Robert_Williams_(robot_fatality))



法治封面

“自动驾驶”：安全，不安全！？

碰撞前 司机没有任何刹车痕迹







# List of self-driving car fatalities

## Level 2 fatalities [edit]

Level 2 is considered automated driving, but not autonomous driving. A Level 2 driving system expects a driver to be fully aware at any time of the driving and traffic situation and be able to take over any moment.

List of known automated driving system car fatalities (occurring while automated driving-system acknowledged to have been engaged)

Date	Incident no.	Country	City	State/county /province	No. of fatalities	System manufacturer	Vehicle Type	Distance driven by the system at time of incident	Notes
20 January 2016	1	China	Handan	Hebei	1	Tesla (Autopilot)	Model S <sup>[10]</sup>	—	Driver fatality. [11][12]
7 May 2016	2	United States of America (USA)	Williston	Florida	1	Tesla (Autopilot)	Model S <sup>[8]</sup>	130,000,000 mi 210,000,000 km <sup>[13][14]</sup>	Driver fatality. [15][16]
23 March 2018	4	United States of America (USA)	Mountain View	California	1	Tesla (Autopilot)	Model X <sup>[8]</sup>	—	Driver fatality. [17]

## Level 3 fatalities [edit]

A Level 3 autonomous driving system would occasionally expect a driver to take over control.

List of known autonomous car fatalities (occurring while autonomous-system acknowledged to have been engaged)

Date	Incident no.	Country	City	State/county /province	No. of fatalities	System manufacturer	Vehicle Type	Distance driven by the system at time of incident	Notes
18 March 2018	3	United States of America (USA)	Tempe	Arizona	1	Uber	'Refitted Volvo' <sup>[8]</sup>	—	Pedestrian fatality. <sup>[9]</sup>

# Hacking of devices/robots

**CYBERSECURITY INSIGHT**

## Hacking Robots Before Skynet<sup>1</sup>

Cesar Cerudo (@cescer) Chief Technology Officer, IOActive  
Lucas Apa (@lucasapa) Senior Security Consultant, IOActive

### Research Preview

This paper is based on our own research and discovery of critical cybersecurity issues in several robots. While we focus the vendors in addressing the cybersecurity vulnerabilities identified, we want to describe the currently available technology, some of the threats posed by a compromised robot, and the types of cybersecurity issues we discovered. The goal is to make robots more secure and prevent vulnerabilities from being used maliciously by attackers to cause serious harm to businesses, consumers, and their surroundings.

**IOActive.**

<sup>1</sup> [https://en.wikipedia.org/wiki/Skynet\\_\(Terminator\)](https://en.wikipedia.org/wiki/Skynet_(Terminator))

merlos in insecurity, IoT, Security ① 17/10/2017 366 Words

## Hack a BT Low Energy (BLE) butt plug



Few weeks ago I bought a [Bluetooth Low Energy \(BLE\)](#) butt plug to test the (in)security of BLE protocol.

This caught my attention after [researchers told us](#) that a lot of sex toys use this protocol to [allow remote control](#) that is insecure by design.

The great [Simone evilsocket Margaritelli](#) wrote a BLE scanner called [BLEAH](#) ([get it on github](#)) and a [wonderful post](#) on how to use it to hack BLE devices. I strongly suggest you to read the post before moving on.

Everything started as a joke between me and Simone before going to [HackInBo](#) (the best free security event in Italy) when Simone said to me “if you buy it we [pentest it in Bologna](#)“.

I bought it 😊 and during HackInBo we hacked the butt plug using [BLEAH](#) (Lovense Hush model). Simone deserves all the credits for this, I just learned from him.

Follow ...

[source] <https://scubarda.com/2017/10/17/hacking-a-bt-low-energy-ble-butt-plug>

# Who's to blame?

## WHO IS TO BLAME WHEN A ROBOT COMMITS A CRIME?

Legal scholar Gabriel Hallevy offers 3 potential scenarios:

### 1. Perpetrator via another



It is determined that the robot was used by a programmer, or other person, to intentionally commit a crime.

**OUTCOME:** The programmer is guilty with intent; the robot is innocent.

### 2. Natural probable cause



It is determined that the crime wasn't intentional, but should have been "reasonably foreseen" by the programmer/manufacturer.

**OUTCOME:** The programmer or manufacturer is charged with negligence; the robot is innocent.

### 3. Direct liability



It is determined that a robot acted with will, volition, or control in committing a crime.

**OUTCOME:** The robot is found guilty of both the action (crime), and the intent to commit it.

SOURCE: "When Robots Kill: AI under Criminal Law" (Hallevy)

theHUSTLE

[more info] <https://www.amazon.com/When-Robots-Kill-Artificial-Intelligence/dp/1555538053>

# Impact on the workforce

**MIT  
Technology  
Review**

Log in / Create an account   Search  ↗

Topics+   The Download   Magazine   Events   More+

[Subscribe](#)

---



A View from **Kai-Fu Lee**

---

## Tech companies should stop pretending AI won't destroy jobs

No matter what anyone tells you, we're not ready for the massive societal upheavals on the way.

February 21, 2018

---

I took an Uber to an artificial-intelligence conference at MIT one recent morning, and the driver asked me how long it would take for autonomous vehicles to take away his job. I told him it would happen in about 15 to 20 years. He breathed a sigh of relief. "Well, I'll be retired by then," he said.

Good thing we weren't in China. If a driver there had asked, I would have had to tell him he'd lose his job in about 10 years—maybe 15 if he was lucky.

That might sound surprising, given that the US is, and has been, in the lead in AI research. But China is catching up—if it hasn't already—and that rivalry, with one nation playing off the other, guarantees that AI is

[source] <https://www.technologyreview.com/s/610298/tech-companies-should-stop-pretending-ai-wont-destroy-jobs>

# Impact on the scientific community

The screenshot shows a WordPress blog post titled "Some Thoughts on a Mysterious Universe" by Mohammed AlQuraishi. The post is dated December 9, 2018, and has 120 comments. The title of the post is "AlphaFold @ CASP13: “What just happened?”". The content discusses the biennial assessment of protein structure prediction methods (CASP13) and the significance of DeepMind's result. It also touches on the methodology, sociology, and general excitement of the meeting. A table of contents is provided at the bottom of the post. The post is part of a blog with navigation links for Home, Cities, About Blog, and About Me.

Some Thoughts on a Mysterious Universe

by Mohammed AlQuraishi

December 9, 2018

120

## AlphaFold @ CASP13: “What just happened?”

I just came back from [CASP13](#), the biennial assessment of protein structure prediction methods (I previously blogged about [CASP10](#).) I participated in a panel on deep learning methods in protein structure prediction, as well as a predictor (more on that later.) If you keep tabs on science news, you may have heard that [DeepMind’s debut went rather well](#). So well in fact that not only did they take first place, but put a [comfortable distance](#) between them and the second place predictor ([the Zhang group](#)) in the free modeling (FM) category, which focuses on modeling novel protein folds. Is the news real or overhyped? What is AlphaFold’s key methodological advance, and does it represent a fundamentally new approach? Is DeepMind forthcoming in sharing the details? And what was the community’s reaction? I will summarize my thoughts on these questions and more below. At the end I will also briefly discuss how RGNs, my [end-to-end differentiable](#) model for structure prediction, did on CASP13.

“What just happened?” was a question put to me in exactly these words by at least one researcher at CASP, and a sentiment expressed by most academics I spoke with. As one myself, I shared it going in and throughout the meeting. In fact I went into CASP13 feeling melancholy (the raw results were out two days prior), although my mood lifted during the meeting due to the general excitement and quality of discussions, and as my tribal reflexes gave way to a cooler and more rational assessment of the value of scientific progress.

This will be a long post. I will start with the science: the significance of DeepMind’s result, their methodology, and how it relates to existing methods. Then I will discuss the sociology: how people reacted, why we did so, what this means for the academic discipline of protein structure prediction (and life science companies), and how I think we ought to move forward. After what I hope is an exposition of general interest, I will briefly discuss how RGNs performed at CASP13. Spoiler alert: not very well, partly because the value of co-evolutionary information increased substantially in this CASP relative to prior ones, and partly because I could not submit the original submissions unaltered owing to technical problems.

For the sake of making this post easier to navigate, below is a table of contents.

Follow ...

[source] <https://moalquraishi.wordpress.com/2018/12/09/alphafold-casp13-what-just-happened>

# Impact on the scientific community

Science & Environment

## AAAS: Machine learning 'causing science crisis'

By Pallab Ghosh  
Science correspondent, BBC News, Washington

© 16 February 2019



AAAS meeting



Astronomy is one of the many areas of science in which machine learning is used to make discoveries

[source]

<https://www.bbc.com/news/science-environment-47267081>

REUTERS

Business Markets World Politics TV More

SCIENCE NEWS MARCH 28, 2012 / 7:10 PM / 7 YEARS AGO

## In cancer science, many "discoveries" don't hold up

Sharon Begley

8 MIN READ



NEW YORK (Reuters) - A former researcher at Amgen Inc has found that many basic studies on cancer — a high proportion of them from university labs — are unreliable, with grim consequences for producing new medicines in the future.



A laboratory researcher in a file photo. REUTERS/Sebastian Derungs

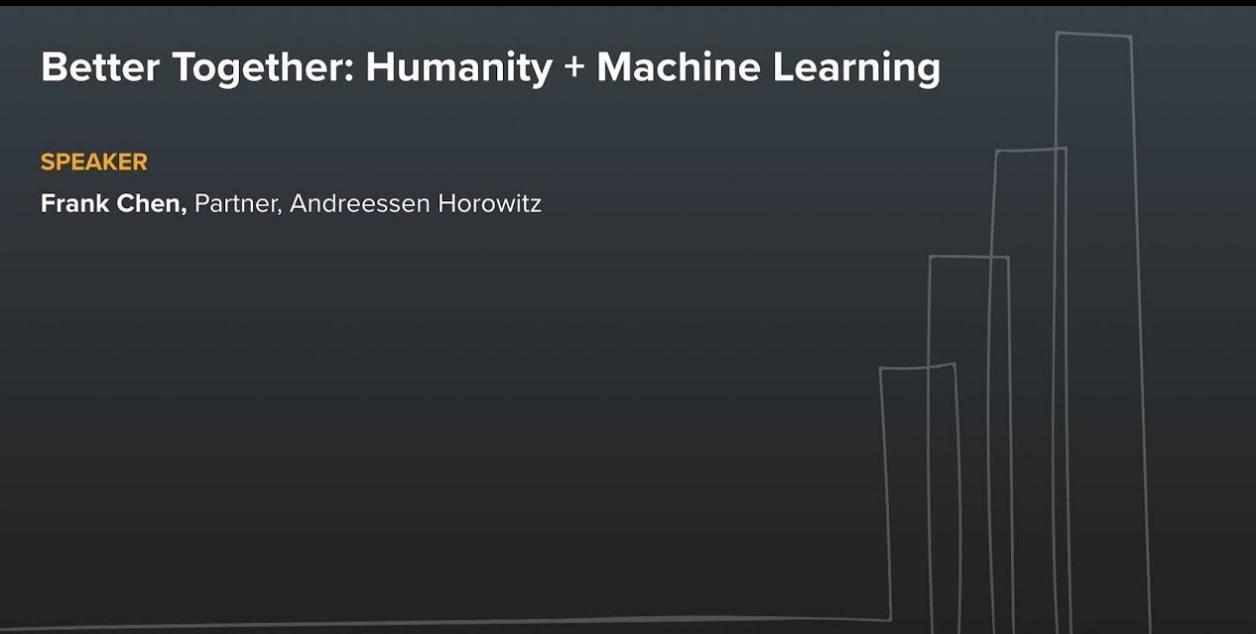
[source]

<https://www.reuters.com/article/us-science-cancer/in-cancer-science-many-discoveries-dont-hold-up-idUSBRE82R12P20120328>

# Better Together: Humanity + Machine Learning

**Better Together: Humanity + Machine Learning**

**SPEAKER**  
**Frank Chen**, Partner, Andreessen Horowitz



a16zSummit

# Sophia said it would 'destroy humans'

BUSINESS  
INSIDER  
UK

TECH

f t in BI Intelligence  
Sign in Edition

Meet the first-ever robot citizen – a humanoid named Sophia that once said it would 'destroy humans'

Chris Weller  
Oct 27, 2017, 6:41 PM 37,783

FACEBOOK LINKEDIN TWITTER EMAIL PRINT

Sophia the robot might not have a heart or brain, but it does have Saudi Arabian citizenship.

As of October 25, Sophia is the first robot in history to be a full citizen of a country.

Sophia was developed by Hanson Robotics, led by AI developer David Hanson. It spoke at this year's Future Investment Initiative, held in the Saudi Arabian capital of Riyadh.

Sophia once said it would "destroy humans," but this time around the robot spoke about its desire to live peacefully among humans.

Here's what the robot is all about.

Sophia the robot is the first robot in history to be granted full citizenship of a country. Graham Flanagan/Business Insider

View As: One Page Slides



[source] <https://www.businessinsider.com/meet-the-first-robot-citizen-sophia-animatronic-humanoid-2017-10>

[video] [https://www.youtube.com/watch?v=W0\\_DPi0PmF0](https://www.youtube.com/watch?v=W0_DPi0PmF0)

# Fake news



[source] <https://www.youtube.com/watch?v=cQ54GDm1eL0>

# Fake news: deepfakes

The screenshot shows a web page from the Brookings Institution's TechTank blog. At the top, there is a navigation bar with links to 'ABOUT US', 'EXPERTS', 'EVENTS', 'THE BROOKINGS PRESS', 'BROOKINGS EXEC ED', 'SUPPORT BROOKINGS', 'CAREERS', and a search bar. Below the navigation bar, there is a header with the word 'BROOKINGS' and a menu with categories like 'CITIES & REGIONS', 'GLOBAL DEVELOPMENT', 'INTERNATIONAL AFFAIRS', 'U.S. ECONOMY', 'U.S. POLITICS & GOVERNMENT', and 'MORE'. The main content area features a large image showing a person in profile looking at a large screen displaying multiple video feeds of a man speaking. Below this image, there is a red banner with the title 'TECHTANK Artificial intelligence, deepfakes, and the uncertain future of truth' and the author's name 'John Villasenor' followed by the date 'Thursday, February 14, 2019'. At the bottom of the page, there are social media sharing icons (Facebook, Twitter, LinkedIn, Print, Email) and a definition of 'Deepfakes'.

**TECHTANK**

**Artificial intelligence, deepfakes, and the uncertain future of truth**

John Villasenor - Thursday, February 14, 2019

**Deepfakes** are videos that have been constructed to make a person appear to say or do something that they never said or did. With artificial intelligence-based methods for creating deepfakes becoming increasingly sophisticated and accessible, deepfakes are raising a set of challenging policy, technology, and legal issues.

[source] <https://www.brookings.edu/blog/techtank/2019/02/14/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth>

# Fake news: deepfakes/faceswap

The screenshot shows the GitHub repository page for `deepfakes / faceswap`. The repository has 390 commits, 3 branches, 0 releases, 48 contributors, and is licensed under GPL-3.0. The main commit list includes:

File/Commit	Description	Date
<code>torzdf dlib-cnn rotation scale bugfix</code>	Latest commit @186577 7 days ago	
<code>.github</code>	Update issue templates	16 days ago
<code>lib</code>	Fix png icons for older tkinter versions	22 days ago
<code>plugins</code>	dlib-cnn rotation scale bugfix	7 days ago
<code>scripts</code>	Extract - bugfix: Skip non-loaded image	26 days ago
<code>tools</code>	Fix duration for ffmpeg extract	8 days ago
<code>.dockerrcignore</code>	Clearer requirements for each platform (#183)	a year ago
<code>.gitignore</code>	extractor fixes	3 months ago
<code>Dockerfile.cpu</code>	Fix bug in docker image tf:tf:latest-gpu-py3 (#400)	9 months ago
<code>Dockerfile.gpu</code>	Fix bug in docker image tf:tf:latest-gpu-py3 (#400)	9 months ago
<code>INSTALL.md</code>	Update README to reflect Tensorflow restrictions. (#597)	16 days ago
<code>LICENSE</code>	Add GNU General Public License v3.0	a year ago
<code>README.md</code>	Remove training data from README.md	a month ago
<code>USAGE.md</code>	Update USAGE.md	9 months ago
<code>faceswap.py</code>	Gui v3.0b (#436)	8 months ago
<code>requirements.txt</code>	Fix numpy version (numpy 1.16.0 has memory leaks)	24 days ago
<code>setup.cfg</code>	dlib-cnn: Inconsistent image size bugfix	3 months ago
<code>setup.py</code>	Cuda path check bugfix in setup.py	a month ago
<code>tools.py</code>	Implement Alignments tool and other minor fixes (#473)	6 months ago
<code>README.md</code>		

[source] <https://github.com/deepfakes/faceswap>

# Software for automatic fake news detection

			<p>Kontakt</p> <p><b>Silke Wiesemann</b> Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE Fraunhoferstraße 20 53343 Wachtberg, Deutschland Telefon +49 228 9435-103 <a href="#">→ E-Mail senden</a> <a href="#">→ www.fkie.fraunhofer.de</a></p>
		<p>Machine Learning: Social Media-Daten auswerten</p> <p><b>Software für die automatisierte Erkennung von Fake News</b></p> <p>Forschung Kompakt / 1.2.2019</p> <p>Erfundene Nachrichten, verdrehte Fakten – Fake News verbreiten sich rasant im Netz und werden oft unbedacht geteilt, vor allem in den Sozialen Medien. Fraunhofer-Forscherinnen und -Forscher haben ein System entwickelt, das Social Media-Daten automatisiert auswertet und bewusst gestreute Falschmeldungen und Desinformationen gezielt herausfiltert. Das Tool nutzt dafür sowohl inhaltliche als auch Metadaten, wobei es die Klassifikation mit Machine Learning erworben hat und diese in Interaktion mit dem Nutzer während der Anwendung verbessert.</p>  <p>© Fraunhofer FKIE</p> <p>Um Fake News zu erkennen, bewertet das Tool des Fraunhofer FKIE nicht nur Texte, sondern beachtet auch Metadaten in die Analyse ein.</p> <p>sowie mit solchen Texten, die der Nutzer als Fake News klassifiziert. Mithilfe dieser Lernsets wird das System trainiert. Um Falschmeldungen herauszufiltern, wenden die Forscherinnen und Forscher «Machine Learning»-Verfahren an, die automatisiert nach bestimmten Merkmalen in den Texten und den Metadaten suchen. Das können beispielsweise in einem politischen Kontext auf semantischer Ebene Formulierungen und Wortkombinationen sein, die sich wieder im alltäglichen Sprachgebrauch noch in der journalistischen Berichterstattung finden wie »die aktuelle Bundeskanzlerin«. Zu den Merkmalen zählen auch sprachliche Fehler. Dies ist insbesondere dann der Fall, wenn der Autor für die Formulierung der Fake News Deutsch und nicht seine eigene Muttersprache verwendet. Dann deuten etwa falsche Gedankenstriche, Orthografie-, Deklinations- oder Satzbaufehler darauf hin, dass eine Meldung eine Fake News sein könnte. Unangemessene Ausdrücke oder umständliche Formulierungen können ein weiteres Indiz sein.</p> <p>»Füttern wir unser Tool mit vielen Merkmalen, so sucht sich das Verfahren selbstlernend diejenigen heraus, die funktionieren. Entscheidend ist auch, welche «Machine Learning»-Verfahren dann die besten Ergebnisse liefern. Das ist sehr aufwändig, da man die diversen Algorithmen mit unterschiedlichen Kombinationen von Merkmalen durchrechnen lassen muss«, so Schade.</p>	

[source] <https://www.fraunhofer.de/de/presse/presseinformationen/2019/februar/software-fuer-die-automatisierte-erkennung-von-fake-news.html>

# Repercussions

The New York Times

## *Wielding Rocks and Knives, Arizonans Attack Self-Driving Cars*



A Waymo autonomous vehicle in Chandler, Ariz., where the driverless cars have been attacked by residents on several occasions. Caitlin O'Hara for The New York Times

By Simon Romero

Dec. 31, 2018

f t m b 1203

[source] <https://www.nytimes.com/2018/12/31/us/waymo-self-driving-cars-arizona-attacks.html>

# Repercussions

**HLN** NIEUWS SPORT SHOWBIZ nina REGIO VIDEO  
BINNENLAND BUITENLAND

Nieuws > Binnenland

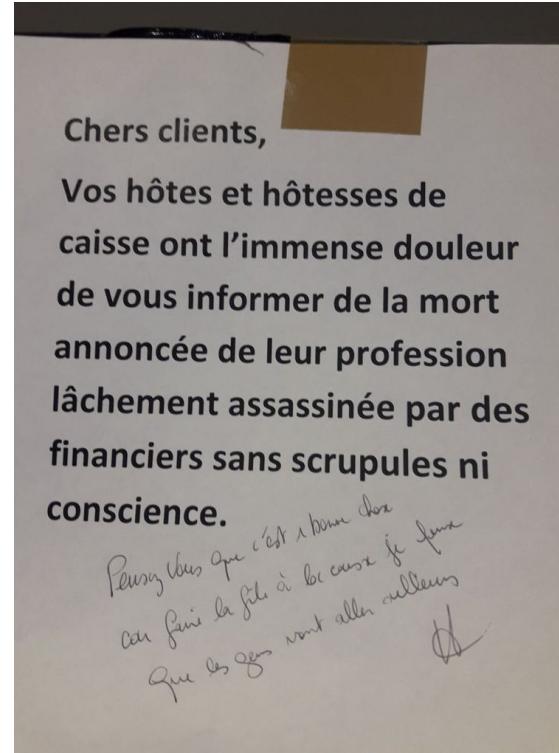
## Kassiersters Carrefour saboteren selfscankassa's: "Ze betekenen onze dood"

FT | 15 februari 2018 | 07u52 | Bron: La Province & Het Nieuwsblad

[DEEL](#) [TWITTER](#) [REACTIES](#) 44 REACTIES



© Borgerhoff - Carrefour herhaalt dat de klassieke kassa niet zal verdwijnen.



# Repercussions

## Kassiersters vechten tegen windmolens

15/02/2018 om 15:51 door **Johan Rasking**



Zelfscanning lukt hier niet in Bergen. Foto: Peter Malaise

**Met een boycotactie van de selfscankassa's proberen de werkneemsters bij Carrefour in Bergen hun job te reden. Dat zal niet lukken. Technologische vooruitgang laat zich niet indijken.**

[source] [http://www.standaard.be/cnt/dmf20180215\\_03358699](http://www.standaard.be/cnt/dmf20180215_03358699)

## Kassiersters saboteren selfscankassa's in Carrefour: "Ons beroep gaat dood"

Vandaag om 08:49 door Kristof Simoens



Foto: Peter Malaise



4  
shares

**"Boycot de selfscankassa's! Want die betekenen de dood van de kassiersters." Oproepen als deze doen op sociale media de ronde in de nasleep van de grote reorganisatie bij Carrefour. In de hypermarkt van Bergen hebben actievoerders de bewuste kassa's zelfs al 'ingepakt'.**

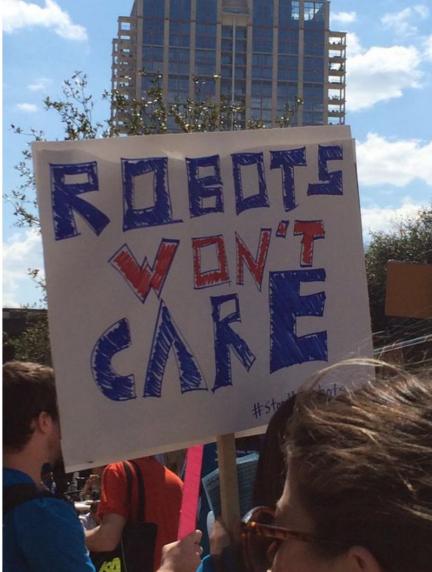
"Beste klanten. Uw hosts en hostessen melden u met veel pijn in het hart de aangekondigde dood van hun beroep, laf vermoord door financiers zonder scruples of geweten." Het plan van Carrefour om in ons land zó'n 1.200 banen te schrappen en gelijktijdig fors te investeren in zogeheten self-checkoutkassa's - kassa's die het zonder Carrefour-medewerker kunnen stellen - blijft de gemoeders beroeren. "Op termijn dreigen alle kassiersters hun job te verliezen", zegt Geoffrey Manfroy van de socialistische bediendevakbond. "Hier, in de hypermarkt van Mons, staan 22 jobs op de helling. Uit protest hebben de kassiersters dit weekend alle selfscankassa's in plastic ingepakt." Volgens de vakbondsman hebben de klanten alle begrip voor de actie. "Meer zelfs, ze steunen het protest actief en sommigen roepen op om de selfscan te boycotten, niet alleen bij ons, ook in andere supermarkten."

# Anti robot rally @ SXSW 2015: publicity stunt

**Anti-Robot Protest Held At SXSW**

Ron Miller @ron\_miller 4 years ago

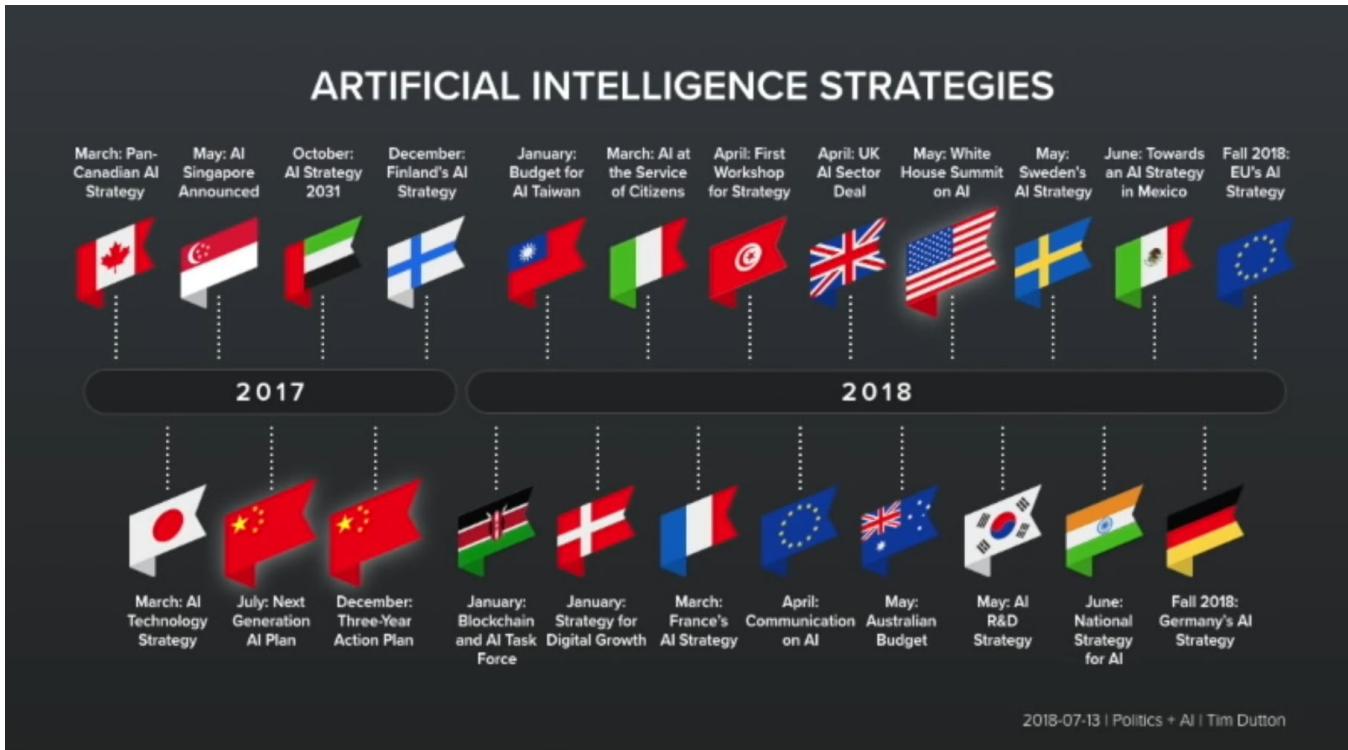
Comment



A small group of protesters held signs and handed out t-shirts to protest robots today at South by Southwest in Austin, Texas.

[info] <https://techcrunch.com/2015/03/14/anti-robot-protest-held-at-sxsw>

# (Geo)politics



[source] <https://a16z.com/2019/02/08/better-together-humanity-machine-learning-chen-summit>

# (Geo)politics

The screenshot shows a white page with a navigation bar at the top featuring the White House logo, links for ECONOMY, NATIONAL SECURITY, BUDGET, IMMIGRATION, and THE OPIOID CRISIS, and a search icon. Below the navigation is a section titled "EXECUTIVE ORDERS" with a red underline. The main title of the document is "Executive Order on Maintaining American Leadership in Artificial Intelligence". Below the title, it says "INFRASTRUCTURE & TECHNOLOGY" and "Issued on: February 11, 2019". There are three gold stars below this information. On the left side, there is a "SHARE" button with icons for Facebook, Twitter, and Email, and a link to "ALL NEWS". The main text begins with "By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:". A detailed section titled "Section 1. Policy and Principles" follows, discussing AI's role in driving economic growth, enhancing national security, and improving quality of life, emphasizing the United States' leadership in AI research and development.

SHARE: [f](#) [t](#) [e](#)

ALL NEWS

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. Policy and Principles. Artificial Intelligence (AI) promises to drive growth of the United States economy, enhance our economic and national security, and improve our quality of life. The United States is the world leader in AI research and development (R&D) and deployment. Continued American leadership in AI is of paramount importance to maintaining the economic and national security of the United States and to shaping the global evolution of AI in a manner consistent with our Nation's values, policies, and priorities. The Federal Government plays an important role in facilitating AI R&D, promoting the trust of

[source] <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence>

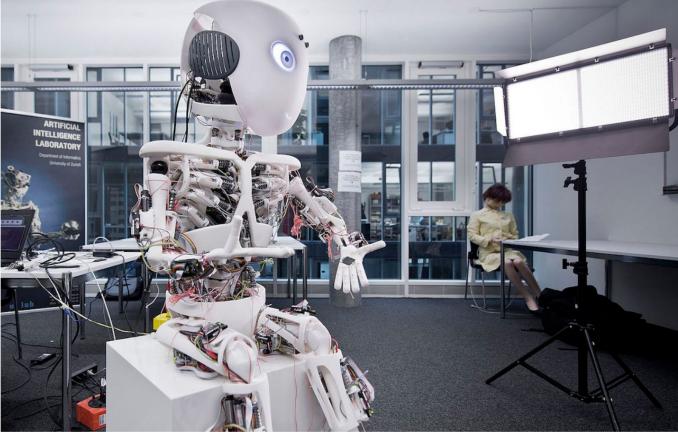
# (Geo)politics

SCIENCE / TECH, TOP STORIES

Published on February 14, 2019 — comments 26

## Understanding China's Confucian Edge in the Global AI Race

written by Craig Smith



[source] <https://quillette.com/2019/02/14/understanding-chinas-confucian-edge-in-the-global-ai-race>

# (Geo)politics



[source] <https://a16z.com/2019/02/08/better-together-humanity-machine-learning-chen-summit>

# Open source

The screenshot shows the GitHub organization page for Facebook Research. The header includes links for Pull requests, Issues, Marketplace, and Explore. The main area features a repository card for "Facebook Research" with a blue icon and the name "Facebook Research". Below it are cards for several repositories:

- PartAI**: A framework for training and evaluating AI models on a variety of openly available dialogue datasets. (Python, 4.162 stars, 1962 forks, MIT license, last updated 10 hours ago)
- pytext**: A natural language modeling framework based on PyTorch. (Python, 4.665 stars, 432 forks, 2 issues need help, last updated 14 hours ago)
- StarSpace**: Learning embeddings for classification, retrieval and ranking. (C++, 2.379 stars, 329 forks, MIT license, last updated a day ago)
- nevergrad**: A Python toolbox for performing gradient-free optimization. (Python, 1.369 stars, 150 forks, MIT license, last updated 2 days ago)
- visdom**: A flexible tool for creating, organizing, and sharing visualizations of live, rich-data. Supports Torch and Numpy. (Python, 0.357 stars, 677 forks, 12 issues need help, last updated 2 days ago)

On the right, there's a sidebar with "Top languages" (Python, C++, C, Lua, Jupyter Notebook, HTML) and "Most used topics" (cpp, deep-learning, machine-learning, neural-network).

[source] <https://github.com/facebookresearch>

The screenshot shows the GitHub organization page for Google AI Research. The header includes links for Pull requests, Issues, Marketplace, and Explore. The main area features a repository card for "Google AI Research" with a multi-colored icon and the name "Google AI Research". Below it are cards for several repositories:

- google-research**: Google AI Research. (Python, 1.14k stars, 163 forks, last updated 10 hours ago)
- computation-thru-dynamics**: Understanding computation in artificial and biological recurrent networks through the lens of dynamical systems. (Python, 2.2k stars, 2 forks, Apache-2.0 license, last updated 10 hours ago)
- bert**: TensorFlow code and pre-trained models for BERT. (tf, natural-language-processing, google, tensorflow). (Python, 12.110 stars, 2.519 forks, Apache-2.0 license, last updated a day ago)
- planet**: Deep Planning Network: Control from pixels by latent planning with learned dynamics. (Python, 227 stars, 28 forks, Apache-2.0 license, last updated a day ago)

On the right, there's a sidebar with "Top languages" (Python, Jupyter Notebook) and "Most used topics" (machine-learning, natural-language-processing, research, tensorflow). A note at the bottom states: "This organization has no public members. You must be a member to see who's a part of this organization."

[source] <https://github.com/google-research>

- Brand / Sales
- Recruiting
- Quality / Community

# Open source: ethical limits?

RESEARCH SYSTEMS

ABOUT BLOG

FEBRUARY 14, 2019

## Better Language Models and Their Implications

We've trained a large-scale unsupervised language model which generates coherent paragraphs of text, achieves state-of-the-art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization — all without task-specific training.

[VIEW CODE](#)

[READ PAPER](#)

[READ MORE](#)

[source] <https://blog.openai.com/better-language-models>

# Open source: ethical limits?

## Release Strategy

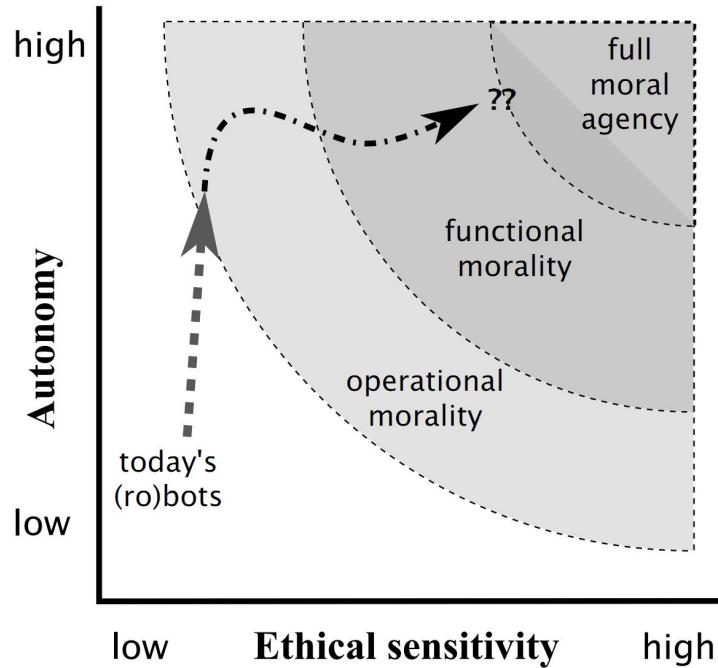
Due to concerns about large language models being used to generate deceptive, biased, or abusive language at scale, we are only releasing a much smaller version of GPT-2 along with sampling code. We are not releasing the dataset, training code, or GPT-2 model weights. Nearly a year ago we wrote in the [OpenAI Charter](#): “we expect that safety and security concerns will reduce our traditional publishing in the future, while increasing the importance of sharing safety, policy, and standards research,” and we see this current work as potentially representing the early beginnings of such concerns, which we expect may grow over time. This decision, as well as our discussion of it, is an experiment: while we are not sure that it is the right decision today, we believe that the AI community will eventually need to tackle the issue of publication norms in a thoughtful way in certain research areas. Other disciplines such as biotechnology and cybersecurity have long had active debates about responsible publication in cases with clear misuse potential, and we hope that our experiment will serve as a case study for more nuanced discussions of model and code release decisions in the AI community.

# WEF: Top 9 ethical issues in artificial intelligence



1. **Unemployment.** What happens after the end of jobs?
2. **Inequality.** How do we distribute the wealth created by machines?
3. **Humanity.** How do machines affect our behaviour and interaction?
4. **Artificial stupidity.** How can we guard against mistakes?
5. **Racist robots.** How do we eliminate AI bias?
6. **Security.** How do we keep AI safe from adversaries?
7. **Evil genies.** How do we protect against unintended consequences?
8. **Singularity.** How do we stay in control of a complex intelligent system?
9. **Robot rights.** How do we define the humane treatment of AI?

# Levels of moral agency



1. **Operational morality:**  
The moral significance of their actions lies entirely in the humans involved in their design and use.
2. **Functional morality:**  
The ability of an Artificial Moral Agent (AMA) to make moral judgments when deciding a course of action without direct instructions from humans.
3. **Full moral agency:**  
Full-fledged sentient robots

# Ethics of artificial intelligence

*The part of the ethics of technology specific to robots and other artificially intelligent beings.*

-- Wikipedia

# Robot ethics & Machine ethics

1. Robot ethics or roboethics:  
The moral behavior of humans as they design, construct, use and treat artificially intelligent beings.
  
2. Machine ethics:  
How to implement moral decision-making in computers and robots?

# Robotics & AI ethical principles round up (1/2)



1. Asimov's three laws of Robotics (1950)
2. Murphy and Wood's three laws of Responsible Robotics (2009)
3. EPSRC Principles of Robotics (2010)
4. Future of Life Institute Asilomar principles for beneficial AI (Jan 2017)

# Robotics & AI ethical principles round up (2/2)

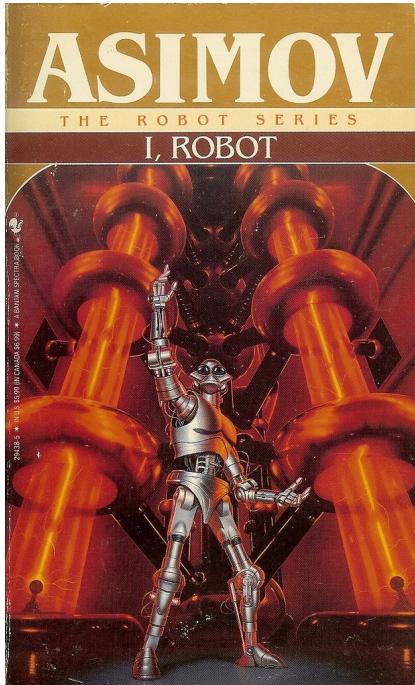


THE  
FUTURE  
SOCIETY



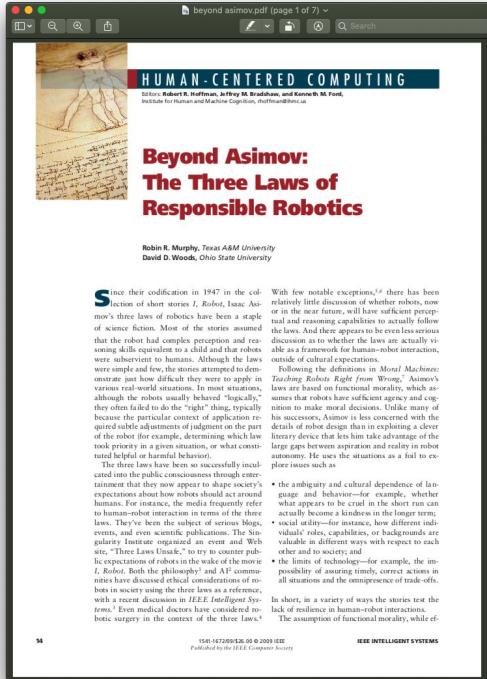
- Future of Life Institute Asilomar principles for beneficial AI (Jan 2017)
- The ACM US Public Policy Council Principles for Algorithmic Transparency and Accountability (Jan 2017)
- Japanese Society for Artificial Intelligence (JSAI) Ethical Guidelines (Feb 2017)
- Draft principles of The Future Society's Science, Law and Society Initiative (Oct 2017)
- Montréal Declaration for Responsible AI draft principles (Nov 2017)
- IEEE General Principles of Ethical Autonomous and Intelligent Systems (Dec 2017)
- UNI Global Union Top 10 Principles for Ethical AI (Dec 2017)

# Asimov's three laws of Robotics (1950)



1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

# Murphy&Wood's 3 laws of Responsible Robotics



1. A human may not deploy a robot without the human-robot work system meeting the highest legal and professional standards of safety and ethics.
2. A robot must respond to humans as appropriate for their roles.
3. A robot must be endowed with sufficient situated autonomy to protect its own existence as long as such protection provides smooth transfer of control which does not conflict with the First and Second Laws.

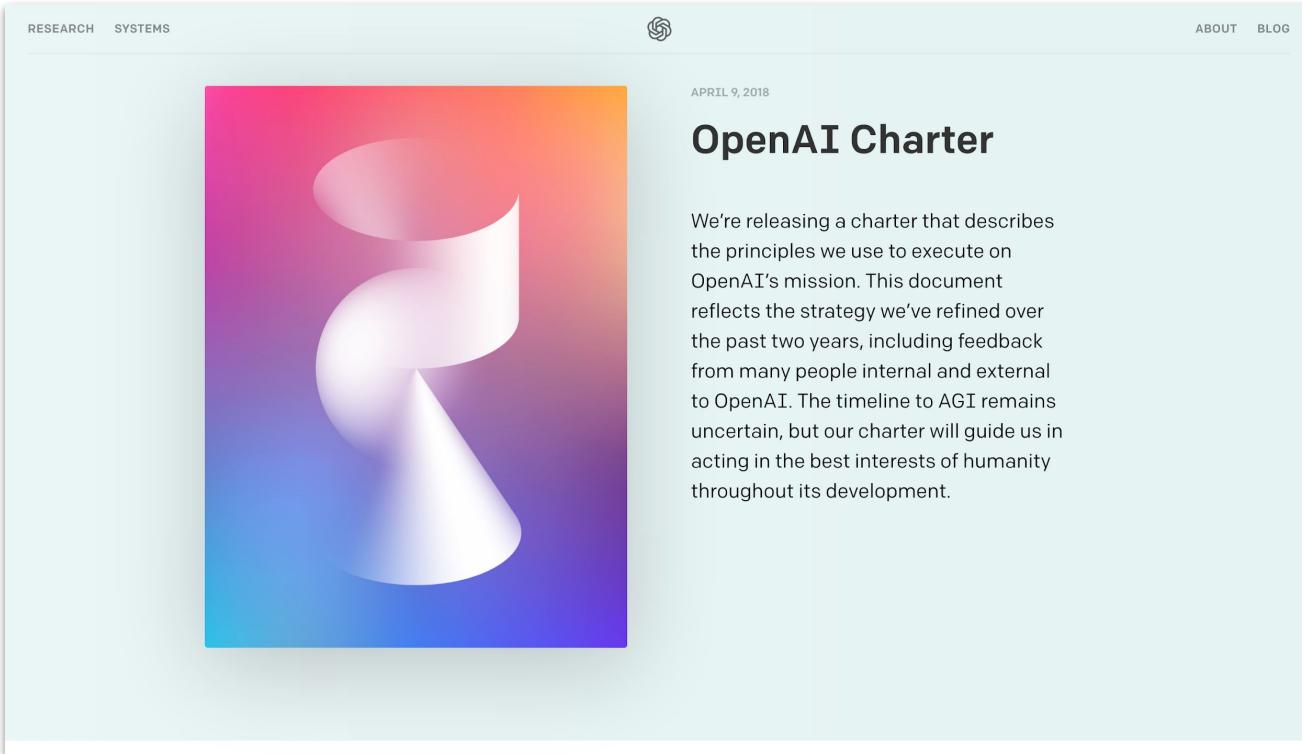
# EPSRC Principles of Robotics (2010)



Engineering and Physical Sciences  
Research Council

1. Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.
2. Humans, not Robots, are responsible agents. Robots should be designed and operated as far as practicable to comply with existing laws, fundamental rights and freedoms, including privacy.
3. Robots are products. They should be designed using processes which assure their safety and security.
4. Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.
5. The person with legal responsibility for a robot should be attributed.

# OpenAI Charter

A screenshot of a web browser displaying the "OpenAI Charter" blog post. The header features the OpenAI logo and navigation links for "RESEARCH", "SYSTEMS", "ABOUT", and "BLOG". The main content area shows a large, colorful graphic of a white ribbon-like shape on a red-to-blue gradient background. To the right, the date "APRIL 9, 2018" is displayed above the title "OpenAI Charter". The text describes the release of a charter outlining principles and strategy for OpenAI's mission, noting feedback from internal and external stakeholders and the uncertainty of AGI timelines.

APRIL 9, 2018

## OpenAI Charter

We're releasing a charter that describes the principles we use to execute on OpenAI's mission. This document reflects the strategy we've refined over the past two years, including feedback from many people internal and external to OpenAI. The timeline to AGI remains uncertain, but our charter will guide us in acting in the best interests of humanity throughout its development.

[source] <https://blog.openai.com/openai-charter>

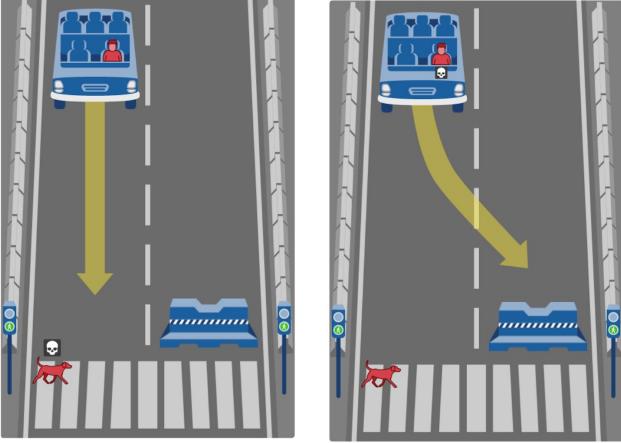
# Moral machine

 MORAL MACHINE

Home Judge Classic Design Browse About Feedback 🇺🇸 En

What should the self-driving car do?

1 / 13



The image displays two side-by-side scenarios for a self-driving car. In both scenarios, the car is heading towards four people tied to the road. In the left scenario, a person is standing near a switch or lever, suggesting a choice can be made. In the right scenario, there is no such lever, implying a different type of decision-making process. Both scenarios include a red dog and a skull, adding to the complexity of the moral choices.

Show Description Show Description

Made by Scalable Cooperation at MIT Media Lab

[source] <http://moralmachine.mit.edu>

# Humans Need Not Apply



# Discussion

