

Ethics of Machine Learning and Data Science - Part I

Week 1: Foundations

Dr Chris Anagnostopoulos

Table of Contents

- 1 Do no harm
- 2 The basic structure of a data science pipeline
- 3 Moral frameworks and codes of conduct
- 4 The five principles

Do no harm

AI is wonderful



Figure 1: <https://deepmind.com/research/case-studies/alphago-the-story-so-far>

AI is wonderful

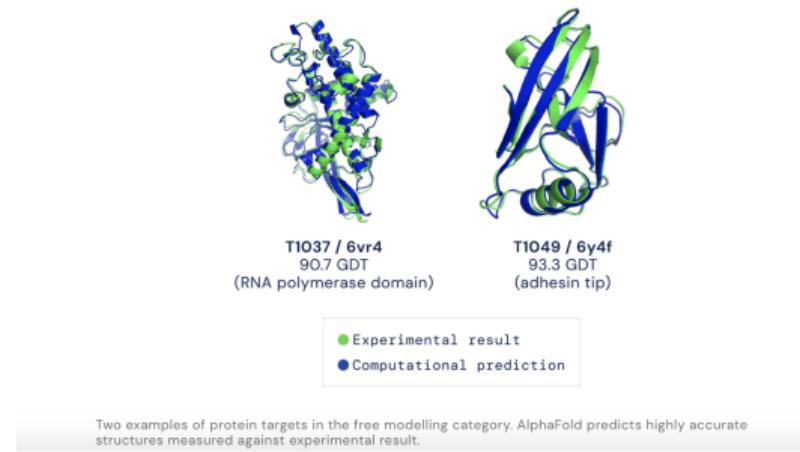


Figure 2: <https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>

AI is wonderful

nature medicine

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature medicine](#) > [letters](#) > [article](#)

Letter | [Published: 07 January 2019](#)

Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network

[Awni Y. Hannun](#)  [Pranav Rajpurkar](#), [Masoumeh Haghpanahi](#), [Geoffrey H. Tison](#), [Codie Bourn](#), [Mintu P. Turakhia](#) & [Andrew Y. Ng](#)

[Nature Medicine](#) **25**, 65–69 (2019) | [Cite this article](#)

45k Accesses | **548** Citations | **367** Altmetric | [Metrics](#)

Figure 3: <https://www.nature.com/articles/s41591-018-0268-3>

AI can cause harm

BBC | [Sign in](#) | [Home](#) | [News](#) | [Sport](#) | [Weather](#) | [iPlayer](#)

NEWS

[Home](#) | [Coronavirus](#) | [Climate](#) | [UK](#) | [World](#) | [Business](#) | [Politics](#) | [Tech](#) | [Science](#) | [Health](#)

[Technology](#)

Uber's self-driving operator charged over fatal crash

16 September 2020

REUTERS

The self-driving Volvo hit a pedestrian at 39mph, despite the presence of a safety driver.

BBC | [Sign in](#) | [Home](#) | [News](#) | [Sport](#) | [Weather](#) | [iPlayer](#)

NEWS

[Home](#) | [Coronavirus](#) | [Climate](#) | [UK](#) | [World](#) | [Business](#) | [Politics](#) | [Tech](#) | [Science](#) | [Health](#)

[Technology](#)

Tesla: Elon Musk suggests Autopilot not to blame for fatal crash

19 April



Two men were killed after a Tesla car crashed into a tree and caught fire in Texas, and police believe there was nobody present in the driver's seat at the time of the accident.

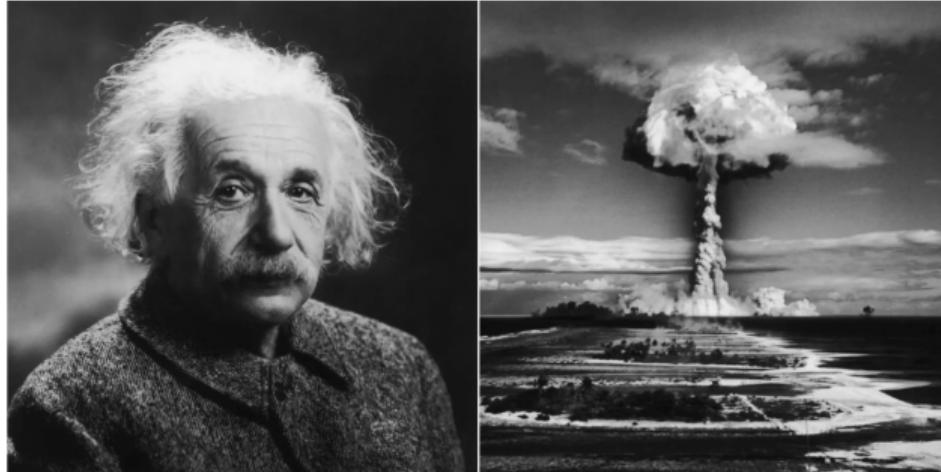
Figure 4: Left: <https://www.bbc.co.uk/news/technology-56799749>, Right: <https://www.bbc.co.uk/news/technology-54175359>

AI can cause harm



Figure 5: Left: <https://www.bbc.co.uk/news/technology-58462511>, Right: <https://www.bbc.co.uk/news/technology-50865437>

Technological adoption relies on public trust



Albert Einstein around 1939 and nuclear explosion in French Polynesia in October 1971. MPI/Getty Images/Michel BARET/Gamma-Rapho via Getty Images

Figure 6: Einstein was initially in favour of the creation of an atomic bomb, but later became vehemently opposed to the proliferation of nuclear arms.

Technological adoption relies on public trust



BBC | Sign In | Home News Sport Weather iPlayer Sound

NEWS

Home | Coronavirus | Climate | UK | World | Business | Politics | Tech | Science | Health | Family & Education

Asia | China | India

Fukushima disaster: What happened at the nuclear plant?

© 2011 BBC

Fukushima nuclear disaster

A screenshot of a BBC news article titled "Fukushima disaster: What happened at the nuclear plant?". The article includes a sub-headline "© 2011 BBC" and a link "Fukushima nuclear disaster". Below the headline is a photograph showing extensive destruction and debris from the 2011 tsunami and nuclear accident in Japan. A caption at the bottom of the image reads "The 2011 tsunami was the most powerful ever recorded in Japan". The BBC navigation bar at the top includes links for Home, News, Sport, Weather, iPlayer, and Sound.

Figure 7: Left: The Chernobyl nuclear reactor (photo by Associated Press), Right: the Fukushima nuclear accident. <https://www.bbc.co.uk/news/world-asia-56252695>

Technological adoption relies on public trust

WIRED

Data Is the New Oil of the Digital Economy

DATA IS THE NEW OIL OF THE DIGITAL ECONOMY

SHARE



SHARE



TWEET



COMMENT
8



EMAIL



- DS and AI are becoming ubiquitous and are gradually transforming the majority of economic, social and political interactions.
- They are too complex to understand without specialised training.
- With understanding, comes responsibility.

Figure 8: The onset of Big Data circa 2014 has ushered a new era for Data Science.

Technological adoption relies on public trust



The collapsed 35W bridge in Minneapolis seen on August 3, 2007.
Many thanks*

Figure 9: Thirteen people were killed and 145 injured during the Interstate 35W bridge collapse in 2007

Many other professions carry a similar burden of responsibility, abide by strict codes of ethics and are legally liable for the outcomes of their work:

- Doctors
- Pharmaceutical companies
- Lawyers
- Safety-critical systems engineers
- ...

The Hippocratic oath

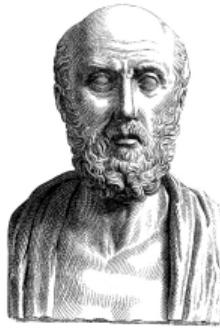


Figure 10: Hippocrates. Unidentified engraver, Public domain, via Wikimedia Commons

First, do no harm

Doctors around the world swear a version of the Hippocratic oath, originating in 400 BCE, and introducing the principles of medical confidentiality and non-maleficence.

Corporate oaths

Google's original internal code of conduct started with the phrase "Don't be evil". This was later rephrased into "*You can make money without being evil.*"

The Hippocratic oath

≡ WIR ED

LONG READS BUSINESS CULTURE GEAR SCIENCE SECURITY VIDEO

SUBSCRIBE NEWSLETTERS



TOM UPCHURCH

BUSINESS 08.04.2018 07:00 AM

To work for society, data scientists need a Hippocratic oath with teeth

Data scientists need to understand the weight of their influence and the limitations of their wisdom, says Cathy O'Neil. The *Weapons of Math Destruction* author lays out her plan for an effective system

Figure 11: Cathy O'Neill in *Weapons of Math Destruction* (O'Neil, 2016) was among the first to make a call for a Hippocratic oath for data scientists.

Doing the right thing is neither obvious nor easy

- Lack of context or poor understanding of the group at-risk.
- Cognitive biases or bad habits.
- Inappropriate incentive structures.
- Inherent trade-offs (moral dilemmas).
- Defeatism.
- Unanticipated consequences.

What is the right thing to do?

We will not attempt to define this here, as this is not a course in moral philosophy.
But I hope it will inspire you to take one, or read a book about it(Naudts, 2019).

Hope for the best, prepare for the worst

Prevalent mindset

- ① Success stories / impact
- ② Access to data
- ③ Do good, not evil
- ④ Someone else will fix it
- ⑤ Unsolvable dilemma
- ⑥ If used properly, it is safe
- ⑦ Technology will fix everything
- ⑧ Only monitor model performance
- ⑨ Keep it between the engineers

Target mindset

- ① Near-misses, anticipate harm
- ② Permission to use data
- ③ Do the best you can, minimise risk
- ④ You hold the bar
- ⑤ Explicit trade-offs
- ⑥ How could this be abused?
- ⑦ Pragmatism and humility
- ⑧ Monitor everything we care about
- ⑨ Engage with broader society

Summary

- With power comes responsibility
- You are best placed to drive positive change
- Anticipate, minimise, communicate
- Human-centred design
- You won't be able to fix everything

Next, we start to become more specific:

- What are we talking about really?
- What progress has been made?
- What principles should guide the design of ethical AI?