Ethics in data science

Getting and Cleaning Data

Data privacy

"The relationship between the collection and dissemination of data, technology, the public expectation of privacy, and the legal and political issues surrounding them."

- Wikipedia

Personally Identifiable Information

"Any information that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual"

- Wikipedia
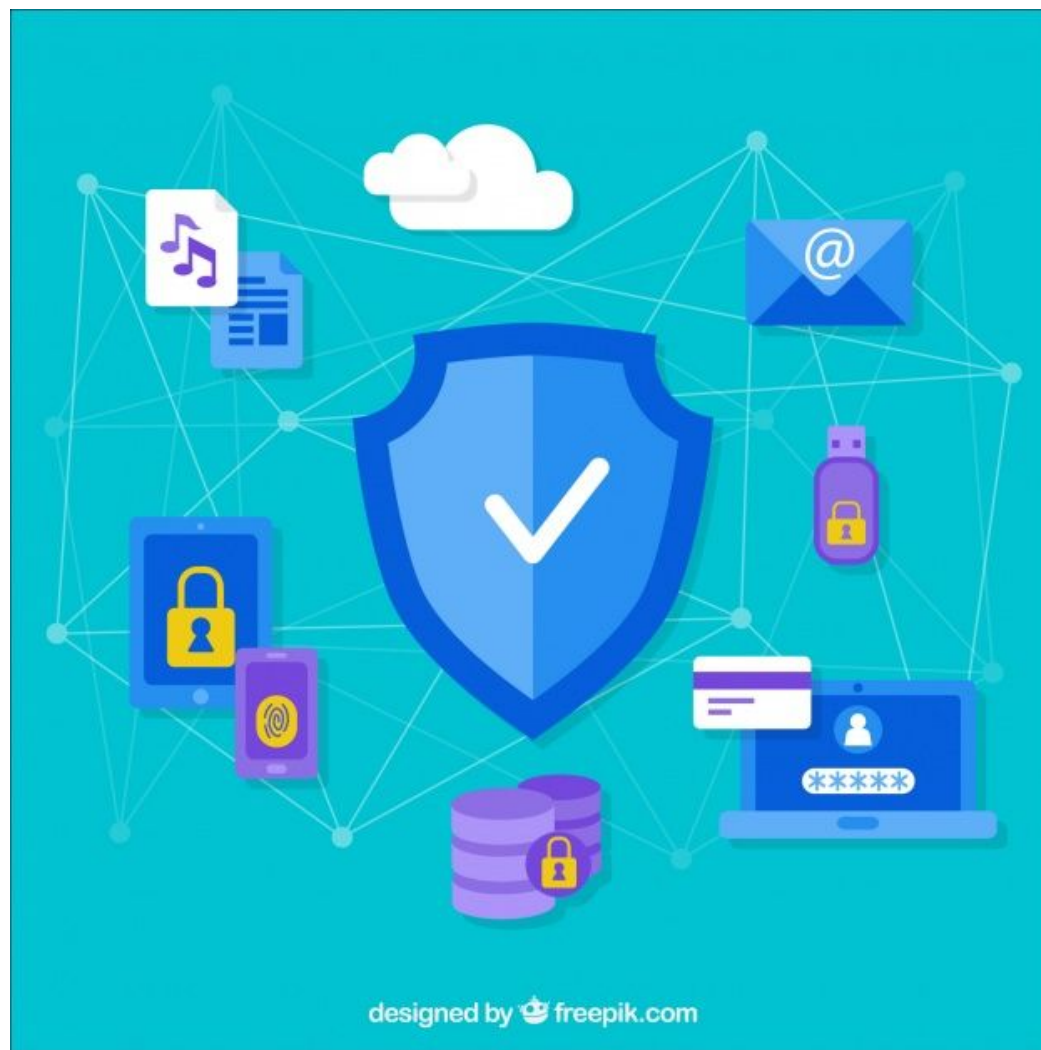
Created by Creativeart - Freepik.com

Data Science projects should **NOT**:

- Increase risk or harm
- Pose threats to individuals
- Make individuals' PII publicly-available or susceptible to theft
- Break any laws
- Share data with other groups/companies without the consent of the individuals
- Save data insecurely

Computer Security

"The protection of computer systems from the theft and damage to their hardware, software or information, as well as from disruption or misdirection of the services they provide."

- Wikipedia

Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

Whistleblower describes how firm linked to former Trump adviser Steve Bannon compiled user data to target American voters

● **'I made Steve Bannon's psychological warfare tool': meet the data war whistleblower**

● **Mark Zuckerberg breaks silence on Cambridge Analytica**

▲ Cambridge Analytica whistleblower: 'We spent $1m harvesting millions of Facebook profiles' – video

https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Equifax traced the source of its massive hack to a preventable software flaw

AP

Ken Sweet and Michael Liedtke, Associated Press

Sep. 14, 2017, 7:55 PM   🔥 17,541

f **FACEBOOK**	in **LINKEDIN**	🐦 **TWITTER**	✉ **EMAIL**	🔗 **COPY LINK**

NEW YORK (AP) — Credit agency Equifax traced the theft of sensitive information about 143 million Americans to a software flaw that could have been fixed well before the burglary occurred, further undermining its credibility as the guardian of personal data that can easily be used for identity theft.

Equifax identified a weakness in an open-source software package called Apache Struts as the technological crack that allowed hackers to heist Social Security

Options traders are betting that Equifax's stock will drop further following last week's announcement of a security breach.   Reuters / Brendan McDermid

TECH • ASHLEY MADISON

What to know about the Ashley Madison hack

Photograph by Philippe Lopez — AFP/Getty Images

By **ROBERT HACKETT** August 26, 2015

Image: Shutterstock

DENMARK

70,000 OkCupid Users Just Had Their Data Published

Joseph Cox
May 12 2016, 1:44pm

Just because data is sort-of public, doesn't mean that it's ethical to collect en masse.

GPS

Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

● **Latest: Strava suggests military users 'opt out' of heatmap as row deepens**
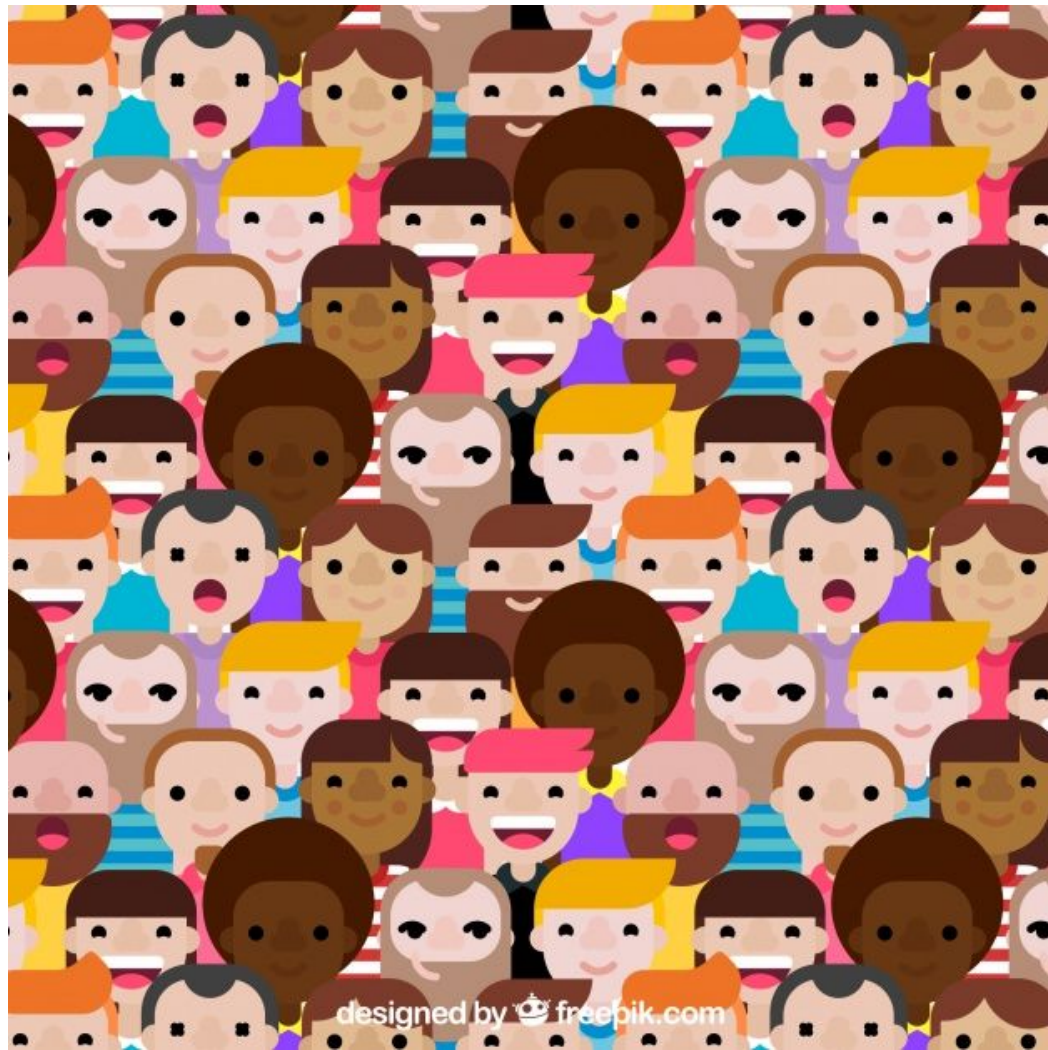
Alex Hern

🐦 @alexhern

Sun 28 Jan 2018 16.51 EST

🖿 This article is over **3 months old**

29,775

▲ A military base in Helmand Province, Afghanistan with route taken by joggers highlighted by Strava. Photograph: Strava Heatmap
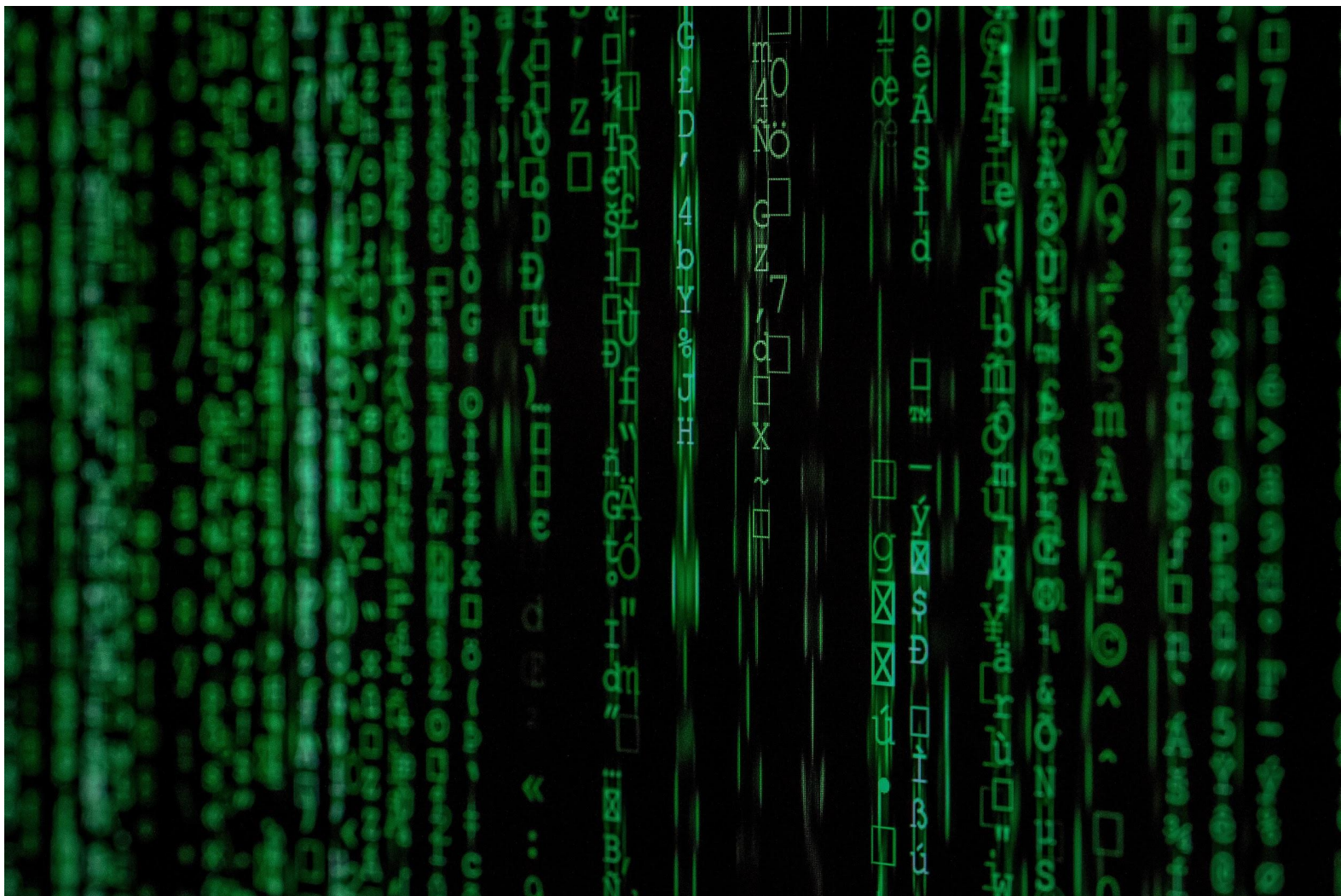
Random sample

Population

Informed Consent

Individuals know:

- how their data will be used

- any risks or benefits

- of any possible consequences

Informed Consent can be withdrawn at any time.

Individual's data must then be removed from dataset.

Before sharing the results of a project, be sure:

1. You've considered bias and ensured your conclusions are fair.

2. You've checked for mistakes.

3. Your work can not be used to harass people, especially minorities, in any way.

4. Your analysis is independent of your opinion about the specific problem you're trying to solve using data.

Data Science Ethics Checklist:

- ❏ Is your team biased?
- ❏ Data collection avoid sampling bias?
- ❏ Are the data biased?
- ❏ Informed consent obtained for all participants?
- ❏ Can one's data be removed on request?
- ❏ Does our analysis discriminate?
- ❏ Is our analysis transparent?
- ❏ What are the negative or unintended consequences of our findings?
- ❏ Do we have a plan to check for biases going forward and update our algorithm?
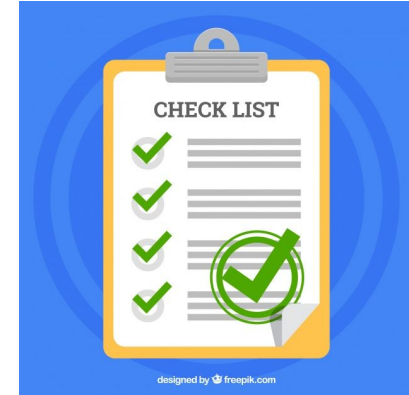
COURTS ARE USING AI TO SENTENCE CRIMINALS. THAT MUST STOP NOW

https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/
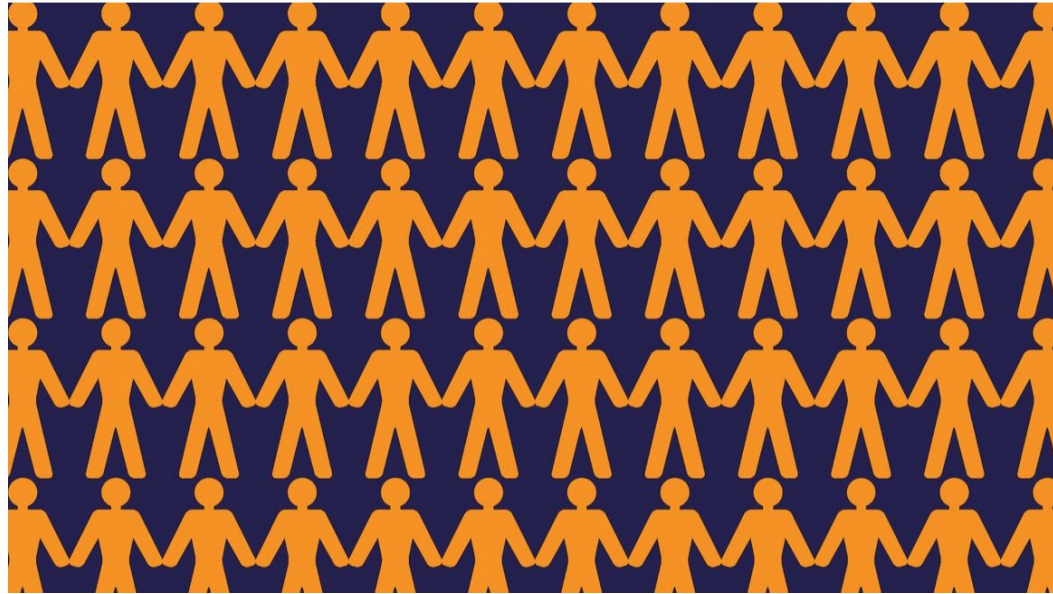
Google Is Helping the Pentagon Build AI for Drones

Kate Conger and Dell Cameron

3/06/18 10:15am • Filed to: GOOGLE ⌄

93.7K   70   3

Graphic: Jim Cooke, Photo: Getty

Amazon Pushes Facial Recognition to Police. Critics See Surveillance Risk.

Amazon promotes its facial recognition technology on the company's website, saying that the service can track people in a video even when their faces are not visible. Amazon

Summarizing:
Ethics in data science

Getting and Cleaning Data