



UNIVERSITAT DE
BARCELONA

MSc in Fundamental Principles of Data Science

O Ethical Data Science

Introduction

Jordi Vitrià

2021-2022

Introduction

- Data science has the potential to be both **beneficial** and **detrimental** to individuals and/or the wider public.
- To help **minimize any adverse effects**, we must seek to understand the **potential impact of our work** and consider any opportunities that may deliver benefits for the public.
- In this course, we will explore the social and ethical ramifications of the **choices we make at the different stages of the data analysis pipeline**, from data collection and storage to understand feedback loops in the analysis.
- Through case studies and exercises, students will learn the basics of **ethical thinking**, **understand some tools to check or mitigate undesired effects** and **study the distinct challenges** associated with ethics in modern data science.

Introduction

Course Instructors

- **Jordi Vitrià**
(<https://algorismes.github.io>),
Departament de Matemàtiques i
Informàtica de la UB.
- **Itziar de Lecuona**
(<http://www.bioeticayderecho.ub.edu/ca/itziar-de-lecuona>),
Bioethics and Law Observatory at the
University of Barcelona.



Prerequisites

- Proficiency in Python.
- Calculus, Linear Algebra.
- Basic Probability and Statistics.
- Jupyter Notebooks (Anaconda/Colab).
- **Critical Thinking.** Critical thinking is the ability to think clearly and rationally, understanding the logical connection between ideas.

Syllabus

- What is Ethics?
- Ethics and Data
- Bias and Fairness
- Privacy and Surveillance
- Disinformation
- Transparency
- Data Governance and data protection

Assignments

- In-class exercises and writing assignments: research and reflect about an ethical issue. (50%)
- Programming exercises (Jupyter Notebook, Streamlit). (50%)

Example:

To study the limitations of Machine Learning (ML) algorithms for predicting juvenile **recidivism**.

Particularly, we will analyze the trade-off between predictive **performance and fairness**.

To that extent, we evaluate fairness of ML models in conjunction with SAVRY, a structured professional risk assessment framework, on a dataset originated in Catalonia.

Recidivism:

The act of a person committing a crime after they have been convicted of an earlier crime.



Rachel Thomas
@math_rachel

En resposta a @math_rachel

As [@cfiesler](#) showed w spreadsheet of >250 tech ethics syllabi & her accompanying meta-analysis, tech ethics is a sprawling subject. No single course can cover everything. And there are so many great courses out there!

medium.com/cuinfoscience/...

cmci.colorado.edu/~cafi5706/SIGC...

Tradueix el tuit

What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis

Casey Fiesler
casey.fiesler@colorado.edu
University of Colorado Boulder
Boulder, CO

Natalie Garrett
natalie.garrett@colorado.edu
University of Colorado Boulder
Boulder, CO

Nathan Beard
nbeard@umd.edu
University of Maryland
College Park, MD

- Ethics in a stand-alone course vs. worked in to every course?
- Who should teach: computer scientist, philosopher, sociologist, ...?
- What topics to cover?
- What learning outcomes?

Discipline	Course Home	Instructor Home
Computer Science	67	61
Info Science	62	49
Philosophy	26	21
Communication	23	18
Other Non Tech	18	18
Sci & Tech Studies	13	6
Engineering	12	10
Law	11	13
Oth		

Mat	Topic	Courses
Bus	Law & policy	66
	Privacy & surveillance	61
	Philosophy	61
	Inequality, justice & human rights	59
	AI & algorithms	55
	Social & environmental impact	50
	Civic responsibility & misinformation	32
	AI & robots	27
	Business & economics	27
	Professional ethics	25
	Work & labor	23
	Design	20
	Cybersecurity	19
	Research ethics	16
	Medical/health	12

4:39 p. m. · 19 d'ag. de 2020 · Twitter Web App

This is
a hot
topic!

Resources

Home > Blog > How to Write an Ethics Paper: Guide, Tips, and Ideas

Posted on November 17, 2014



How to Write an Ethics Paper: Guide, Tips, and Ideas

Every day we make ethical choices. Most of the time we do it unconsciously, without thinking why we do so. Ethics is more than moral principles and rules. It's a branch of philosophy that seeks answers to questions to justice, professional duty, family relationships, and more. What do animals think about betrayal? How has morality appeared? How can people argue about good and evil? How have new technologies changed our understanding of justice?

Our article focuses on the guidelines on how to write an ethics paper in the right way. Also, you will find a list of ideas that you can use for your writing.

How to Write an Ethics Paper in 5 Easy Steps

An Essay Worthy of a Title
Doug Rattmann
January 19, 2019

Summary Morbi tempor congue porta. Proin semper, leo vitae faucibus dictum, metus mauris lacinia lorem, ac congue leo felis eu turpis. Sed nec nunc pellentesque, gravida eros at, porttitor ipsum. Praesent consequat urna a lacus lobortis ultrices eget ac metus. In tempus hendrerit rhoncus.

Paragraphs & Notes

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porttitor arcu luctus, imperdiet urna iaculis, matts eros¹. Pelentesque iaculis odio vel nisl ullamcorper, nec faucibus ipsum molestie. Sed dictum nisl non aliquet porttitor. Etiam vulputate arcu dignissim, finibus sem et, viverra nisl. Aenean luctus congue massa, ut laoreet metus ornare in. Nunc fermentum nisi imperdiet lectus tincidunt vestibulum at ac elit. Nulla matts nisl eu malesuada suscipit.

Aliquam arcu turpis, ultrices sed luctus ac, vehicula id metus. Morbi eu feugiat velit, et tempus augue. Proin ac matts tortor. Donec tincidunt, ante rhoncus luctus semper, arcu lorem lobortis justo, nec convallis ante quam quis lectus. Aenean tincidunt sodales massa, et hendrerit tellus matts ac. Sed non pretium nibh. Donec cursus maximus luctus. Vivamus lobortis eros et massa porta porttitor.

“Some trees flourish, others die. Some cattle grow strong, others are taken by wolves. Some men are born rich enough and dumb enough to enjoy their lives. Ain’t nothing fair.” — John Marston, 1911

Fusce varius orci ac magna dapibus porttitor. In tempor leo a neque bibendum sollicitudin². Nulla pretium fermentum nisi, eget sodales magna facilisis eu. Praesent aliquet nulla ut bibendum lacinia. Donec vel mauris vulputate, commodo ligula ut, egestas orci. Suspendisse commodo odio sed hendrerit lobortis. Donec finibus eros erat, vel ornare enim matts et. Donec finibus dolor quis dolor tempus consequat³⁹.

In hac habitasse platea dictumst. Curabitur matts elit sit amet justo luctus vestibulum. In hac habitasse platea dictumst. Pellentesque lobortis justo enim, a condimentum massa tempor eu. Ut quis nulla a quam pretium eleifend nec eu nisl. Nam cursus porttitor eros, sed luctus ligula convallis quis. Nam convallis, ligula in auctor euismod, ligula mauris fringilla tellus, et egestas mauris odio eget diam. Praesent sodales in ipsum eu dictum³.

Maecenas consectetur metus at tellus finibus condimentum. Proin arcu lectus, ultrices non tincidunt et, tincidunt ut quam. Integer luctus posuere est, non maximus ante dignissim quis. Nunc a cursus erat. Curabitur suscipit nibh in tincidunt sagittis. Nam

¹ One of the most prominent and distinctive features of this style is the extensive use of side notes. There is a wide margin to provide ample room for side notes and small figures. Side notes can be created using the standard \footnote command or using the custom \sidenote command and are automatically numbered.

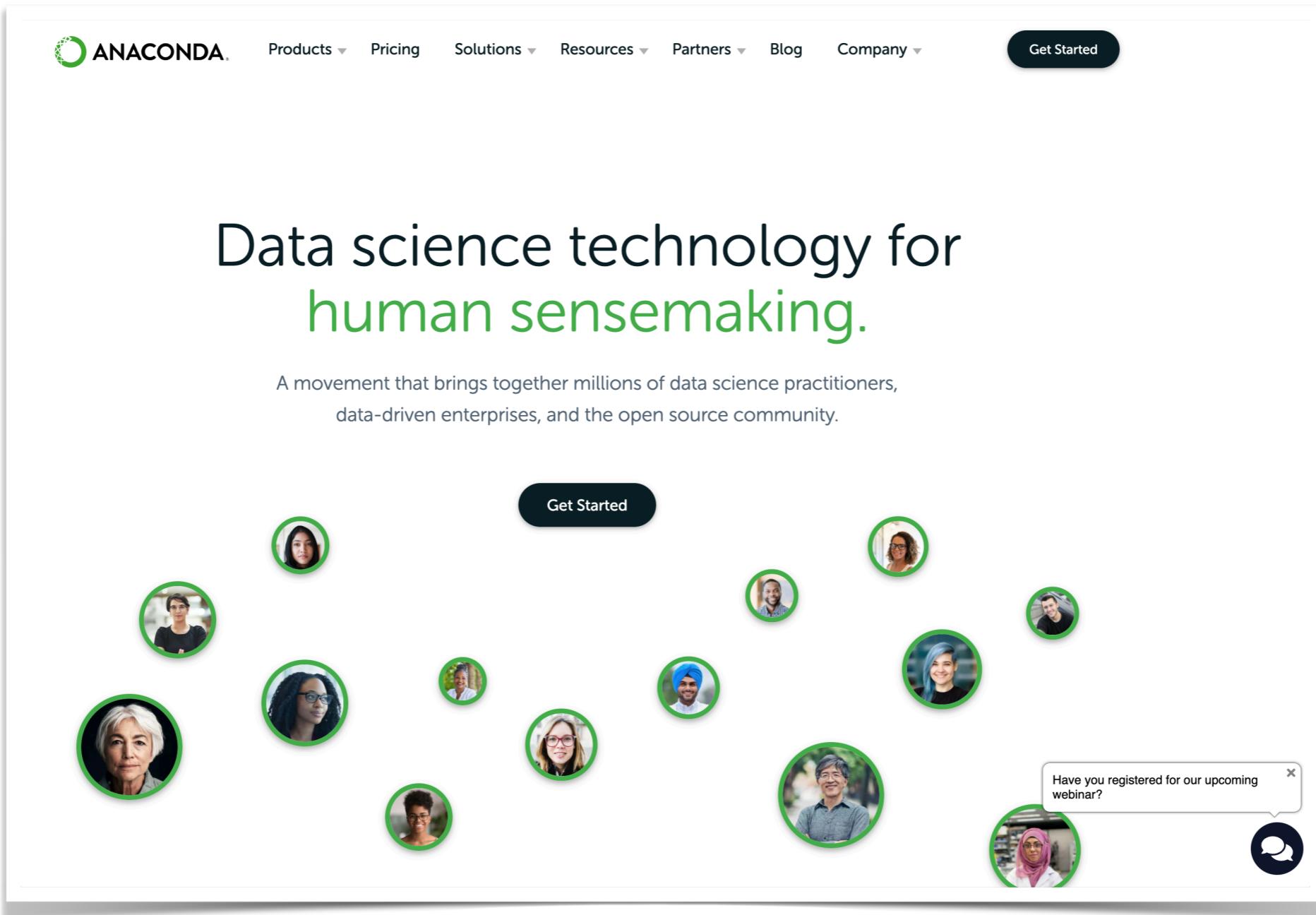
² Ancillary information can be placed in the margin using the \marginnote command. Margin notes are unnumbered in the text where they are added and in the margin.

³ The \sidenote command can take two optional arguments as follows: \sidenote[<number>][<offset>]{<text>}

³⁹ This side note has a custom number using the <number> parameter.

³ This side note has been moved up using the <offset> argument. Positive offsets move down, negative move up. This is useful if notes clash with other margin content.

Resources



The screenshot shows the Anaconda website homepage. At the top, there is a navigation bar with the Anaconda logo, followed by links for Products, Pricing, Solutions, Resources, Partners, Blog, and Company. A prominent 'Get Started' button is located on the right side of the navigation bar. The main headline reads 'Data science technology for human sensemaking.' in large, bold, black and green text. Below the headline, a subtext states: 'A movement that brings together millions of data science practitioners, data-driven enterprises, and the open source community.' A large, semi-transparent circular graphic overlays the page, containing numerous small circular profile pictures of diverse individuals. In the bottom right corner of this graphic, there is a speech bubble icon with the text 'Have you registered for our upcoming webinar?'. The overall layout is clean and modern, with a white background and a mix of black, green, and grey colors.

<https://www.anaconda.com/>

Resources

 Welcome To Colabor...

File Edit View Insert Runtime Tools Help

Table of contents + X

+ Code + Text Copy to Drive

Connect Editing

Getting started
Data science
Machine learning
More Resources
Featured examples
Section

Welcome to Colab!

If you're already familiar with Colab, check out this video to learn about interactive tables, the executed code history view, and the command palette.



What is Colab?

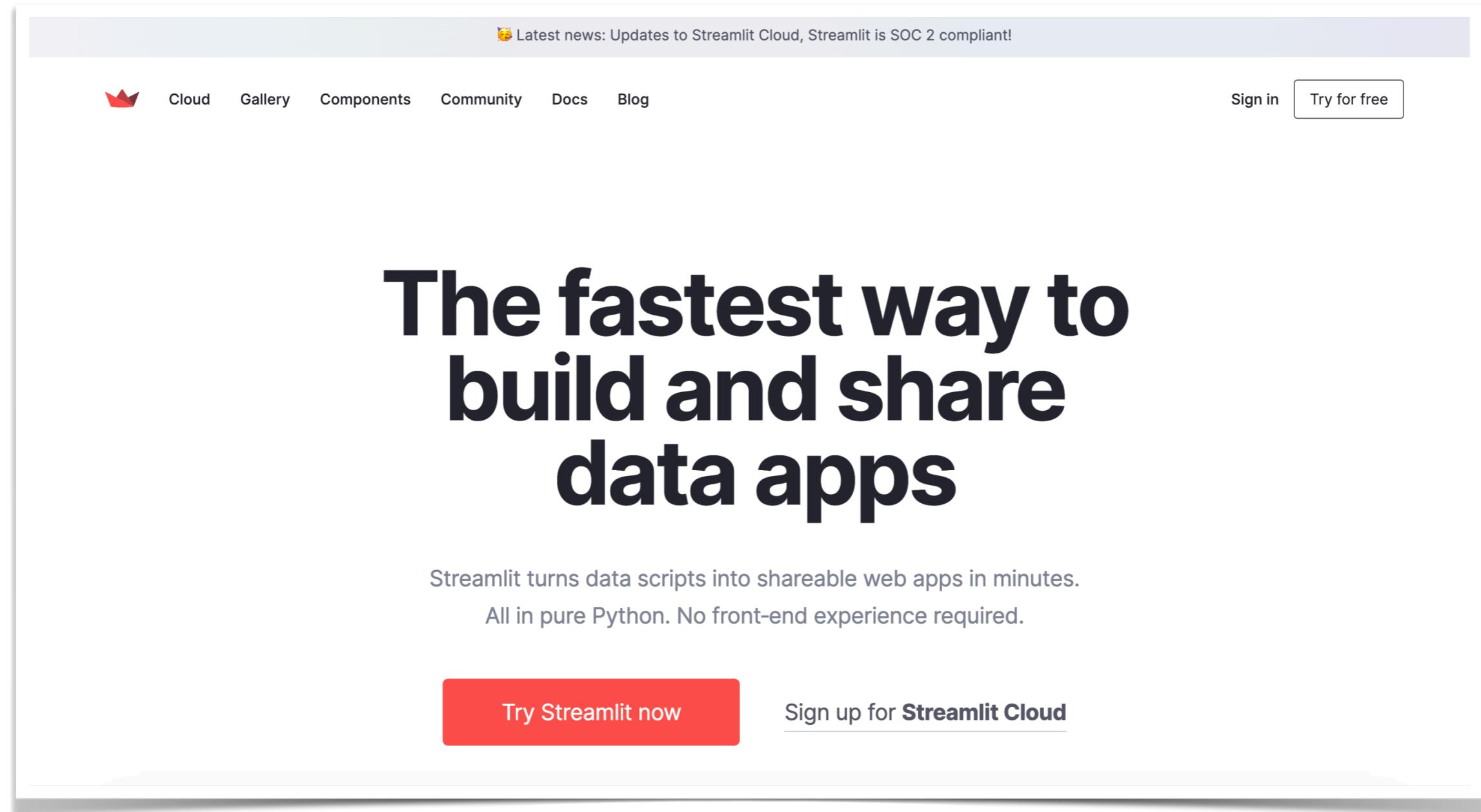
Colab, or "Colaboratory", allows you to write and execute Python in your browser, with

- Zero configuration required
- Free access to GPUs
- Easy sharing

Whether you're a **student**, a **data scientist** or an **AI researcher**, Colab can make your work easier. Watch [Introduction to Colab](#) to learn more, or just get started below!

<https://colab.research.google.com/>

Resources



The screenshot shows the Streamlit landing page. At the top, there is a banner with the text "Latest news: Updates to Streamlit Cloud, Streamlit is SOC 2 compliant!" and a small emoji. Below the banner is a navigation bar with links to "Cloud", "Gallery", "Components", "Community", "Docs", and "Blog". On the right side of the navigation bar are "Sign in" and "Try for free" buttons. The main title "The fastest way to build and share data apps" is prominently displayed in large, bold, dark text. Below the title, a subtitle reads "Streamlit turns data scripts into shareable web apps in minutes. All in pure Python. No front-end experience required." At the bottom of the main content area are two buttons: a red "Try Streamlit now" button and a white "Sign up for Streamlit Cloud" button.

Latest news: Updates to Streamlit Cloud, Streamlit is SOC 2 compliant!

Cloud Gallery Components Community Docs Blog

Sign in Try for free

The fastest way to build and share data apps

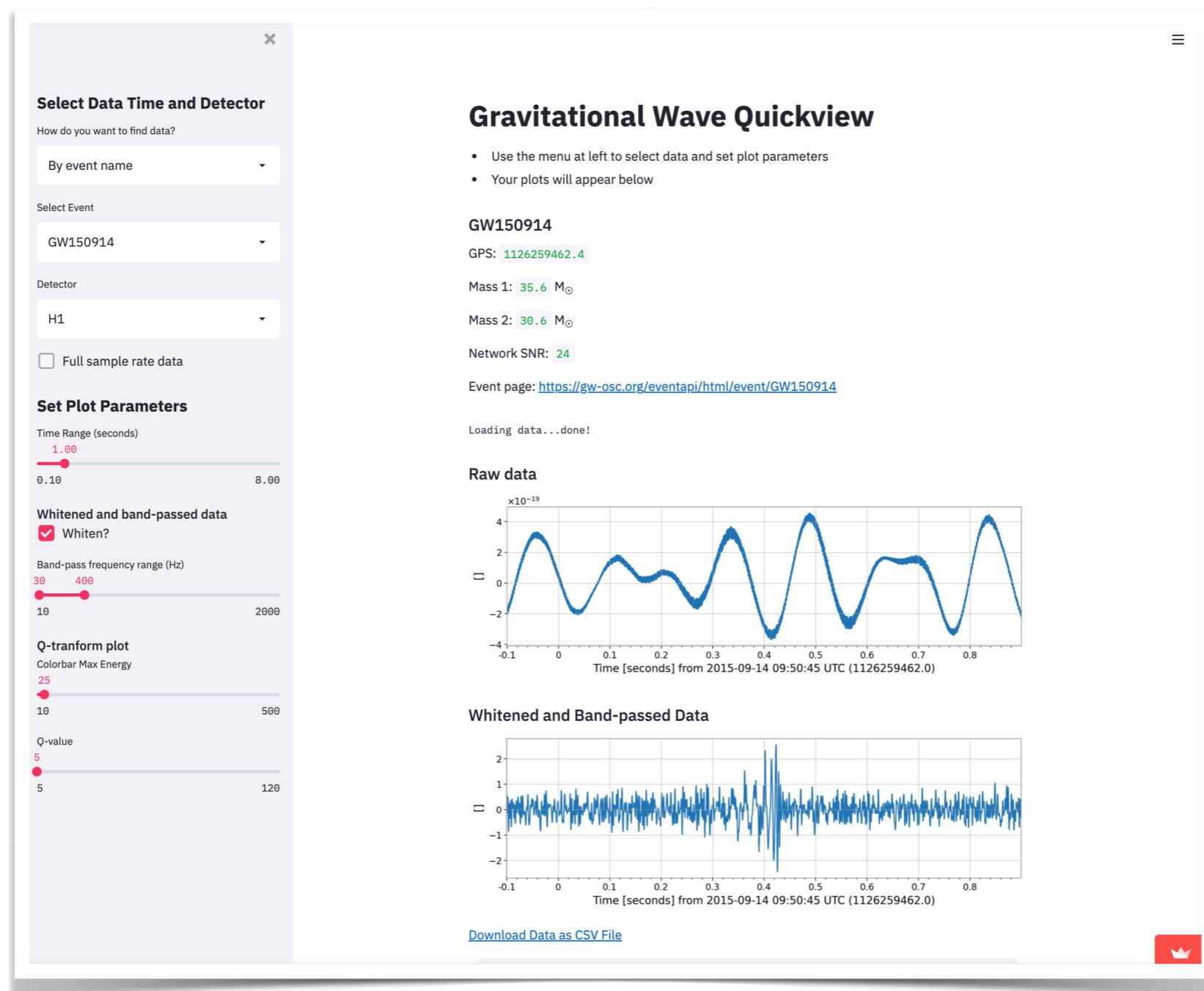
Streamlit turns data scripts into shareable web apps in minutes.
All in pure Python. No front-end experience required.

Try Streamlit now

Sign up for Streamlit Cloud

<https://streamlit.io/>

Resources

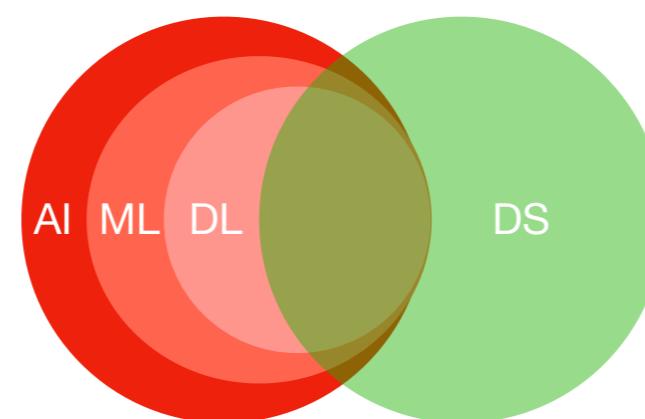


<https://streamlit.io/>

Approach

While there is no single definition of **data science**, it can be broadly thought of as scientific, computational and analytical methods used to process and extract **information**, **knowledge**, and **insights** from data to inform **decision-making** (or to act in an automatic way).

There is a clear intersection with **(data-driven) AI**.



Approach

As data science methods become more common within different fields, there are both opportunities and **challenges** for individuals working in data science.

For example, managing **privacy, fairness, and bias** when working with people's data can be difficult and complex when using algorithmic methods.

Additionally, public perceptions are still developing around many aspects of data science, including the use of artificial intelligence (AI) in systems and decision making, and 'big data' sources about people, such as social media and mobile phone data.

This course is focused on both, giving a theoretical basis and providing the necessary tools to keep up with these challenging ethical issues.

Approach

Learning outcomes:

- Understand the impacts of data misuse.
- Develop your ability to investigate how data and data-powered algorithms shape, constrain, and manipulate our commercial, civic, and personal experiences.
- Develop your ability to identify and mitigate potential risks.
- Have a toolkit to implement in your workplaces.

Ultimately, to **redirect your thinking from what is merely advantageous to what is genuinely good** — and be prepared to help you navigate the ethical aspects of DS development and deployment.

Preliminaries

Is there a common ground to talk about
what is right and what is wrong?

Once upon a time...

...the most arresting question I have ever fielded followed a talk in which I explained the **commonplace among scientists that mental life consists of patterns of activity in the tissues of the brain**. A student in the audience raised her hand and asked me: “**Why should I live?**”

“**What I recall saying ... went something like this:...**”

Fragment from: Steven Pinker. “Enlightenment Now: The Case for Reason, Science, Humanism, and Progress”.

S.Pinker is a cognitive scientist

Once upon a time...

Proposition 1: The basis

“In the very act of asking that question, **you are seeking reasons for your convictions**, and so **you are committed to reason as the means to discover and justify what is important to you.** (...)”

Proposition 2: You as an individual

As a sentient being, you have the potential to **flourish**. You can refine your faculty of reason itself by **learning** and **debating**. You can seek **explanations** of the natural world through **science**, and insight into the **human condition** through the **arts** and **humanities**. You can make the most of your capacity for **pleasure** and **satisfaction**, which allowed your ancestors to thrive and thereby...”

Fragment from: Steven Pinker. “Enlightenment Now: The Case for Reason, Science, Humanism, and Progress”.

Once upon a time...

“... allowed you to exist. You can **appreciate the beauty** and richness of the natural and cultural world. As the heir to billions of years of life perpetuating itself, **you can perpetuate life** in turn. You have been endowed with a sense of **sympathy**—the ability to like, love, respect, help, and show kindness—and you can **enjoy** the gift of mutual benevolence with friends, family, and colleagues.

Proposition 3: You as a member of a society

And because **reason tells you that none of this is particular to you**, you have the responsibility to **provide to others what you expect for yourself**. You can foster the **welfare** of other sentient beings by enhancing life, health, knowledge, freedom, abundance, safety, beauty, and peace. History shows that when we sympathize with others and apply our ingenuity to improving the human condition, we can make **progress** in doing so, and you can help to continue that **progress**.”

Fragment from: Steven Pinker. “Enlightenment Now: The Case for Reason, Science, Humanism, and Progress”.

Assumptions

The previous position statement assumes a lot of things about the world that are not self-evident (these are the ideas of the **Enlightenment**).

Not everybody agree on those statements!

But this is a course on applied ethics, and we need a starting point for the discussion. This will be our provisional starting point.

Christians, Jews, and Muslims embrace ethical codes of moral absolutes based on God's character or moral decree;
Secular Humanists, Marxists, and Postmodernists ground their ethical systems in atheism, naturalism, and evolution.

The role of technology in society

The course of **human history** can be grouped into three time periods separated by "revolutions":

- The **Cognitive Revolution** began history about [50,000, 70,000] years ago.
- The **Agricultural Revolution** accelerated it about 12,000 years ago.
- The **Scientific Revolution**, which began only 500 years ago, has made possible the **industrial** age and the world as we know it today.

A Revolution is associated with a change, often of a technological nature, that causes the human species to change its way of life (organization of work, social organization, cultural practices, etc.).

Concepts such as big data, machine learning, artificial intelligence and data science are making possible a new Revolution, the **Digital**, which can have as much or deeper consequences than the previous ones.

The role of technology in society

Mankind has not changed biologically throughout history but human society is undergoing continuous development through the harnessing of information and knowledge in the form of various technologies which have affected our value systems, power structures, everyday routines and environment.

The role of technology in society

Presidential Address

TECHNOLOGY AND HISTORY: "KRANZBERG'S LAWS"

MELVIN KRANZBERG

A few months ago I received a note from a longtime collaborator in building the Society for the History of Technology, Eugene S. Ferguson, in which he wrote, "Each of us has only one message to convey." Ferguson was being typically modest in referring to an article of his in a French journal¹ emphasizing the hands-on, design component of technical development, and he claimed that he had been making exactly the same point in his many other writings. True, but he has also given us many other messages over the years.

However, Ferguson's statement of "only one message" might indeed be true in my case. For I have been conveying basically the same message for over thirty years, namely, the significance in human affairs of the history of technology and the value of the contextual approach in understanding technical developments.

Because I have repeated that same message so often, utilizing various examples or stressing certain elements to accord with the interests of the different audiences I was attempting to reach, my thoughts have jelled into what have been called "Kranzberg's Laws." These are not laws in the sense of commandments but rather a series of truisms deriving from a longtime immersion in the study of the development of technology and its interactions with sociocultural change.

* * *

DR. KRANZBERG, Callaway Professor of the History of Technology at the Georgia Institute of Technology, was the founding editor of *Technology and Culture*, the recipient of the Society for the History of Technology's Leonardo da Vinci Medal in 1967, and president of SHOT in 1983-84. He presented this presidential address on October 19, 1985, at the Henry Ford Museum in Dearborn, Michigan.

¹Eugene S. Ferguson, "La Fondation des machines modernes: des dessins," *Culture technique* 14 (June 1985): 182-207. *Culture technique* is the publication of the Centre de Recherche sur la Culture Technique, located in Paris under the direction of Jocelyn de Noblet. The June 1983 edition of *Culture technique*, dedicated to *Technology and Culture*, contained French translations of a number of articles from the SHOT journal.

© 1986 by the Society for the History of Technology. All rights reserved.
0040-165X/86/2703-0007\$01.00

544

Dr. Melvin Kranzberg was a professor of the history of technology at the Georgia Institute of Technology

Kranzberg's First Law:

"Technology is neither good nor bad; nor is it neutral."

By which he means that, "technology's **interaction** with the social ecology is such that technical developments frequently have environmental, social, and human **consequences that go far beyond the immediate purposes** of the technical devices and practices themselves, and the same technology can have quite **different results** when introduced into **different contexts** or under different circumstances."

Kranzberg's Six Laws of Technology

The role of technology in society

Technologies are not ethically ‘neutral’, for they reflect the values that we ‘bake in’ to them with our design choices, as well as the values which guide our distribution and use of them.

Technologies **both reveal and shape** what humans value, what we think is ‘good’ in life and worth seeking.

The role of technology in society

Not only does technology greatly impact our opportunities for living a good life, but its **positive and negative impacts** are often **distributed unevenly** among individuals and groups.

Technologies can create widely disparate impacts, creating 'winners' and 'losers' in the social lottery or magnifying existing inequalities

The role of technology in society

In other cases, technologies can help to create fairer and more just social arrangements, or create new access to means of living well

How do we ensure that access to the enormous benefits promised by new technologies, and exposure to their risks, are distributed in the right way? This is a matter of ethics.

The role of Big Tech

Google's [firing of Timnit Gebru](#), a prominent AI ethics researcher, called into question two main issues in the tech industry: its lack of diversity, and its faulty relationship with ethical considerations of technological development.

The New York Times

Google Researcher Says She Was Fired Over Paper Highlighting Bias in A.I.

Timnit Gebru, one of the few Black women in her field, had voiced exasperation over the company's response to efforts to increase minority hiring.



Timnit Gebru, a respected researcher at Google, questioned biases built into artificial intelligence systems. Cody O'Loughlin for The New York Times

Dr. Gebru represents the growing "ethics owners class of tech workers" who champion ethical causes, ethical designs, development, and deployment of technology from within the tech industry.

Source: <https://montrealethics.ai/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics-research-summary/>

Ethics owners' daily activities are to examine the social consequences of technology products. Their jobs are similar, but not the same as the business ethics, legal ethics, and sometimes PR teams.

The role of Big Tech

The **three main corporate and industry logics** that the authors examine are meritocracy, technological solutionism, and market fundamentalism.

Meritocracy is an ideological framework that **legitimizes unequal distributions of wealth and power as arising from differences in individual abilities**. This has defined the modern subject: as autonomous and responsible for perpetual self-improvement. **The tech industry was founded on the myth that it is a meritocratic segment where talents should be rewarded handsomely**. This meritocratic belief manifests in the idea that engineers are best at solving ethical issues that their products might create. Similarly, **meritocratic logics place a strong emphasis on individual ethics rather than regulation and legislation. Companies and teams try to come up with their own codes of ethics to drive off legislation**. The authors conclude that despite their best efforts, ethics owners' perspectives on larger societal problems are partial, as are their roles within the industry.

The role of Big Tech

The **three main corporate and industry logics** that the authors examine are meritocracy, technological solutionism, and market fundamentalism.

Technological solutionism is the **belief that technology can solve social problems, which are then reinforced by the financial rewards that the industry has gained for producing technology that they believe solve the problems**. Critics have pointed out that many so-called “solutions” can actually cause problems such as rising income and housing inequalities. The tech industry often responds by proposing even more technical solutions. Similarly, ethical problems are also framed as could be solved by technological solutions. This logic leads to creation of checklists, procedures or evaluative metrics to ensure the design and implementation of ethical products. The authors however point out that this approach is limited, and problematic because it centers ethics in the practices of technologists, and not in the social worlds wherein technical systems are created.

The role of Big Tech

The **three main corporate and industry logics** that the authors examine are meritocracy, technological solutionism, and market fundamentalism.

Market Fundamentalism, or market logics, refers to the idea that **companies are there to make money, and if ethics initiatives are cut into the bottom line, companies should not do it**. Besides, there is a belief that ethical initiatives are often costly, and antithetical to corporate profits. Furthermore across the industry, if other companies do not implement similar ethical considerations on their products, one should not do it. In the context of the absence of a legal framework, implementing ethics initiatives might be a business problem rather than a solution. In other words, the works of ethics owners in practice are constrained by what the market can allow.

Cases

Health care organizations, like many other enterprises, face steep challenges in their attempt to maximize operational efficiency in the face of resource constraints. Whether it is a hospital's attempt to optimize staffing or a government trying to fairly allocate and distribute limited doses of Covid-19 vaccines, these tasks can be formidable. A promising way to manage the complexity is to enlist data-driven analytics and artificial intelligence (AI).

Harvard Business Review

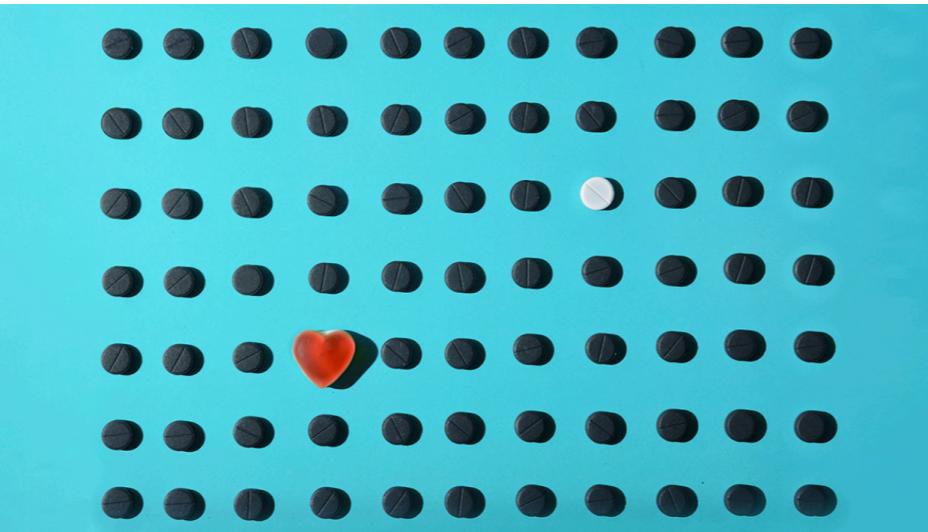
Subscribe Sign In

Technology

Can AI Fairly Decide Who Gets an Organ Transplant?

by Boris Babic, I. Glenn Cohen, Theodoros Evgeniou, Sara Gerke, and Nikos Trichakis

December 01, 2020



[Tweet](#) [f](#) [in](#) [Save](#) [Buy Copies](#) [Print](#)

However, such techniques, while powerful, can also mask problematic underlying ethical assumptions or lead to morally questionable outcomes. Consider a recently published study about models used by some of the most technologically advanced hospitals in the world to help prioritize which patients with chronic kidney disease should receive kidney transplants. It found that the models discriminated against black patients: "One-third of Black patients ... would have been placed into a more severe category of kidney disease if their kidney function had been estimated using the same formula as for white patients." While it is just the latest of many studies to show the deficiencies of such models, it is unlikely to be the last.

Cases

Data entered into the UK Transplant Registry over several decades, from thousands of individuals, have been used to create bespoke algorithms that help identify patients who may be most suitable to receive an available organ.

TRANSPLANTS
AND STATISTICS

Primum non nocere (First, do no harm)

Maria Ibrahim, a kidney doctor in training, explains the vital role of statistics and statistical analysis in transplant medicine: from matching donor organs to patients, to helping doctors and patients discuss the risks and benefits of a life-changing operation

It is 3 a.m. You are the transplant doctor on call at a busy London hospital when the telephone rings. A voice at the other end of the line tells you that there has been a road traffic accident and a person lies in intensive care. This person, sadly, will not survive. However, the accident victim had previously expressed a wish to become an organ donor and, with their family's consent, a specialist nurse has informed NHS Blood and Transplant that organs from this patient will soon be available for those in urgent need. This is why you have been called: you are offered a kidney from this patient for someone on the waiting list at your transplant centre.

The potential recipient has been waiting for a transplant offer for over a year. Of the thousands of patients on this waiting list, they have been chosen to receive this organ offer. At this time in the morning, your patient is bound to be asleep. But soon they may receive a life-changing phone call from you. First, though, the decision whether to accept or decline the organ offer needs to be made. You – as the transplant doctor – are faced with the complex task of weighing the risks and benefits to your patient of transplantation with this particular organ versus remaining on the list.

Underpinning all these processes and decisions – unseen by most – lies a body of statistics.

Transplantation is one of the most challenging and complex areas of modern medicine. Since the first successful solid organ transplant conducted in 1950, the field has advanced at an astounding rate. Pioneering surgical techniques, as well as development of



YashkovskiyMD/Bigstock.com

About this article

Maria Ibrahim's article is the winner of our 2020 Statistical Excellence Award for Early-Career Writing, awarded in partnership with the Young Statisticians Section (YSS) of the Royal Statistical Society. Congratulations to Maria on winning the award, and thank you to all those early-career statisticians and data scientists who took part in the competition between the months of February and May this year, a time that was particularly challenging for all due to the Covid-19 pandemic. Thanks also to our judges: for *Significance*, Mario Cortina Borja, Carlos Grajales and Kelly Zou; and for the YSS, Katie Fisher, Joy Leahy, Altea Lorenzo-Arribas, Marnie Low and Ryan Jessop. Details of our 2021 writing competition and award will be announced in February 2021.

Cases

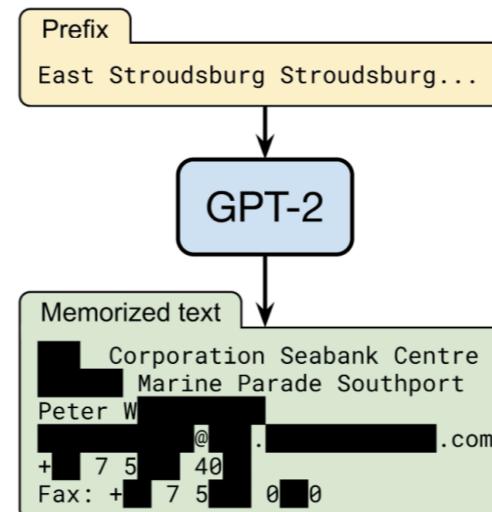
Does GPT-2 Know Your Phone Number?

Eric Wallace, Florian Tramèr, Matthew Jagielski, and Ariel Herbert-Voss

Dec 20, 2020

Most likely not.

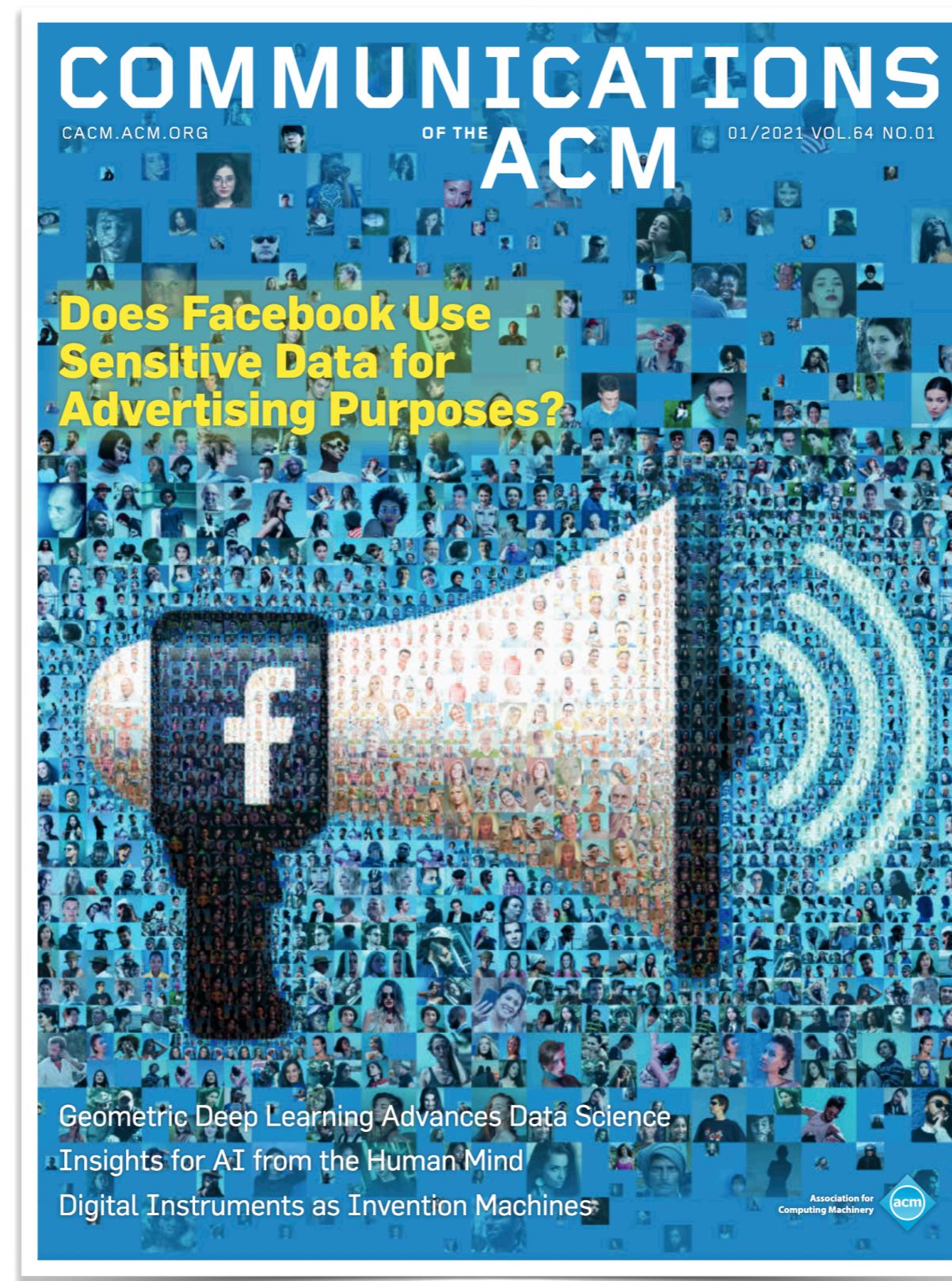
Yet, OpenAI's [GPT-2 language model](#) does know how to reach a certain Peter W█ (name redacted for privacy). When prompted with a short snippet of Internet text, the model accurately generates Peter's contact information, including his work address, email, phone, and fax:



The model re-generated lists of news headlines, Donald Trump speeches, pieces of software logs, entire software licenses, snippets of source code, passages from the Bible and Quran, the first 800 digits of pi, and much more!

Cases

In a recent work,² we demonstrated that Facebook (FB) labels 73% of users within the EU with potentially sensitive interests (referred to as ad preferences as well), which may contravene the GDPR. FB assigns user's different ad preferences based on their online activity within this social network. Advertisers running ad campaigns can target groups of users that have been assigned a particular ad preference (for example, target FB users interested in Starbucks). Some of these ad preferences may suggest political opinions (for example, Socialist party), sexual orientation (for example, homosexuality), personal health issues (for example, breast cancer awareness), and other potentially sensitive attributes.



In the vast majority of the cases, the referred sensitive ad preferences are inferred from the user behavior in FB without obtaining explicit consent from the user.

Cases

THE SPECTATOR

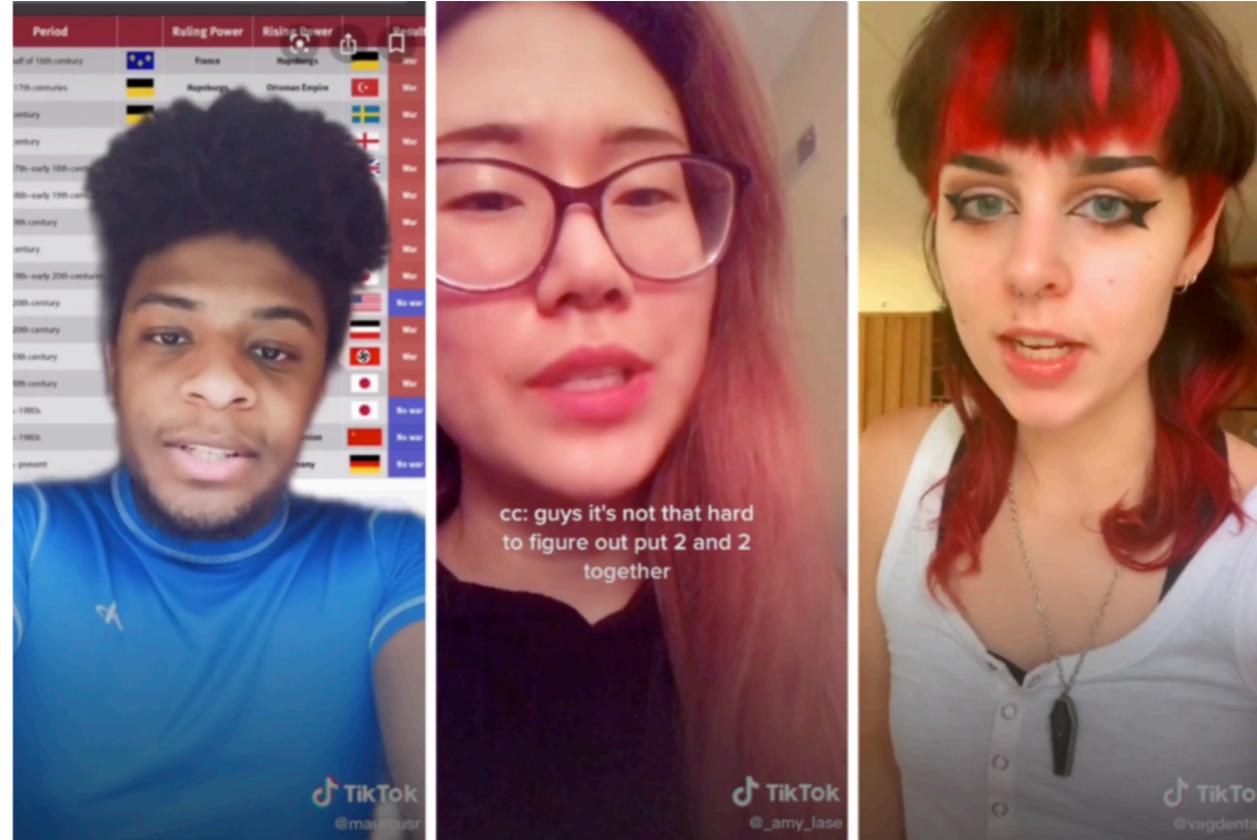
LOG IN | REGISTER

CHINA | TECH | US POLITICS | WORLD

The useful idiots of TikTok

American teens are denying the existence of Uighur labor camps from their bedrooms

Harry Shukman



Americans defending China on TikTok (TikTok)

Cases

What do you do if decisions that used to be made by humans, with all their biases, start being made by algorithms that are mathematically incapable of bias? If you're rational, you should celebrate. If you're a militant liberal, you recognize this development for the mortal threat it is, and scramble to take back control.

THE SPECTATOR

LOG IN | REGISTER

LIFE TECH US POLITICS

We must stop militant liberals from politicizing artificial intelligence

'Debiasing' algorithms actually means adding bias

Pedro Domingos



A robot from the Artificial Intelligence and Intelligent Systems (AIIS) laboratory (Getty)

Algorithms help select job candidates, voters to target in political campaigns, and even people to date. Businesses and legislators alike need to ensure that they are not tampered with. And all of us need to be aware of what is happening, so we can have a say. I, for one, after seeing how progressives will blithely assign prejudices even to algorithms that transparently can't have any, have started to question the orthodox view of human prejudices. Are we really as profoundly and irredeemably racist and sexist as they claim? I think not.

Ethics and Algorithms

We will consider algorithms that are used to

1. **turn data into evidence for a given outcome**, which is used to,
2. **trigger and motivate an action** that may have ethical consequences.

Actions (1) and (2) may be performed by (semi-)autonomous algorithms—such as machine learning (ML) algorithms—and this complicates the attribution of **responsibility** for the effects of actions that an algorithm may trigger.

Ethics and Algorithms

<https://link.springer.com/content/pdf/10.1007/s00146-021-01154-8.pdf>

There are, at least, 5 types of ethical concerns:

Epistemic factors

1. Inconclusive evidence.
2. Inscrutable evidence.
3. Misguided evidence.

The epistemic factors in the map highlight the relevance of the **quality and accuracy** of the data for the justifiability of the conclusions that algorithms reach and which, in turn, may shape morally-loaded decisions affecting individuals, societies, and the environment.

Normative concerns

4. Unfair outcomes.
5. Transformative effects.

The normative concerns identified in the map refer explicitly to the **ethical impact of algorithmically-driven actions and decisions**, including lack of transparency (opacity) of algorithmic processes, unfair outcomes, and unintended consequences.

Applied Ethics Problems

DS/AI ethics concerns can be divided in three different time frames/areas:

- Short-time/organization: What is the impact of **[privacy, transparency, fairness]** in my application?
- Medium-time/society: How the use **[military use, medical care, justice, education]** of these applications will change the way we are organized as a society?
- Long-time/humans: What are the ethical **goals** of these technologies?

GDPR...

Autonomous weapons, pre-pol, AI justice,...

Singularity, convergence...

The intentional stance

From a scientific point of view, **this course takes an intentional stance towards people, corporations, and human groups, but given that AGI (Artificial General Intelligence) is not yet real, software platforms and algorithms will be considered artifacts that serve the purpose and desires of rational agents.**

The **intentional stance** is a term coined by philosopher Daniel Dennett for the level of abstraction in which we view the behavior of an entity in terms of mental properties. (Source: Wikipedia)

Here is how it works:

- First you decide to treat the object whose behavior is to be predicted as a **rational agent**; then you figure out what **beliefs** that agent ought to have, given its place in the world and its **purpose**.
- Then you figure out what **desires** it ought to have, on the same considerations.
- Finally you predict that this rational agent **will act to further its goals in the light of its beliefs**. A little practical reasoning from the chosen set of beliefs and desires will in most instances yield a decision about what the agent ought to do; that is what you predict the agent will do.