UNIVERSITAT DE BARCELONA

Introduction + DS in Context

Jordi Vitrià

Introduction

- Data science has the potential to be both **beneficial** and **detrimental** to individuals and/or the wider public.
- To help **eliminate/mitigate 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 **causal thinking, 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.
- **Critical Thinking.**

Critical thinking is the ability to think clearly and rationally, understanding the logical connection between ideas.

Grading

- The subject will be evaluated through a combination of both an exam (50%) and practical assignments (50%).
  - The exam will test the students' theoretical understanding of the material covered in class and will likely include **short answer questions, and/or essay questions**.
  - The **practical assignments/case studies**, on the other hand, will give students the opportunity to apply what they have learned in class to real-world scenarios and will be used to evaluate their practical skills and abilities.

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

Example:

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

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.

Syllabus

- DS in Context
- What is Ethics?
- Ethics and Data
- Bias and Fairness
- Privacy and Surveillance
- Transparency and explainability
- Data Governance and data protection

Calendar (tentative)

- Feb 15 | Introduction + Data Science in Context.
- Feb 22 | Ethical Foundations I
- Mar 01 | Ethical Foundations II
- Mar 08 | Bias and Discrimination I
- Mar 15 | Bias and Discrimination II
- Mar 22 | Causality
- Mar 29 | Bias and Discrimination III
- Apr 12 | Bias and Discrimination IV
- Apr 19 | Transparency and Explainability
- Apr 26 | no lecture (Matefest/Infofest day)
- May 3 | Privacy or the problem of data agency
- May 10 | Ethics and data protection in a data-driven society (Itziar de Lecuona)
- May 17 | Data governance (Itziar de Lecuona)
- May 24 | Data protection impact assessment methodologies: data cycle and risk assessment (Itziar de Lecuona)

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 Marylan
College Park, MD

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		
Bus		

- 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?

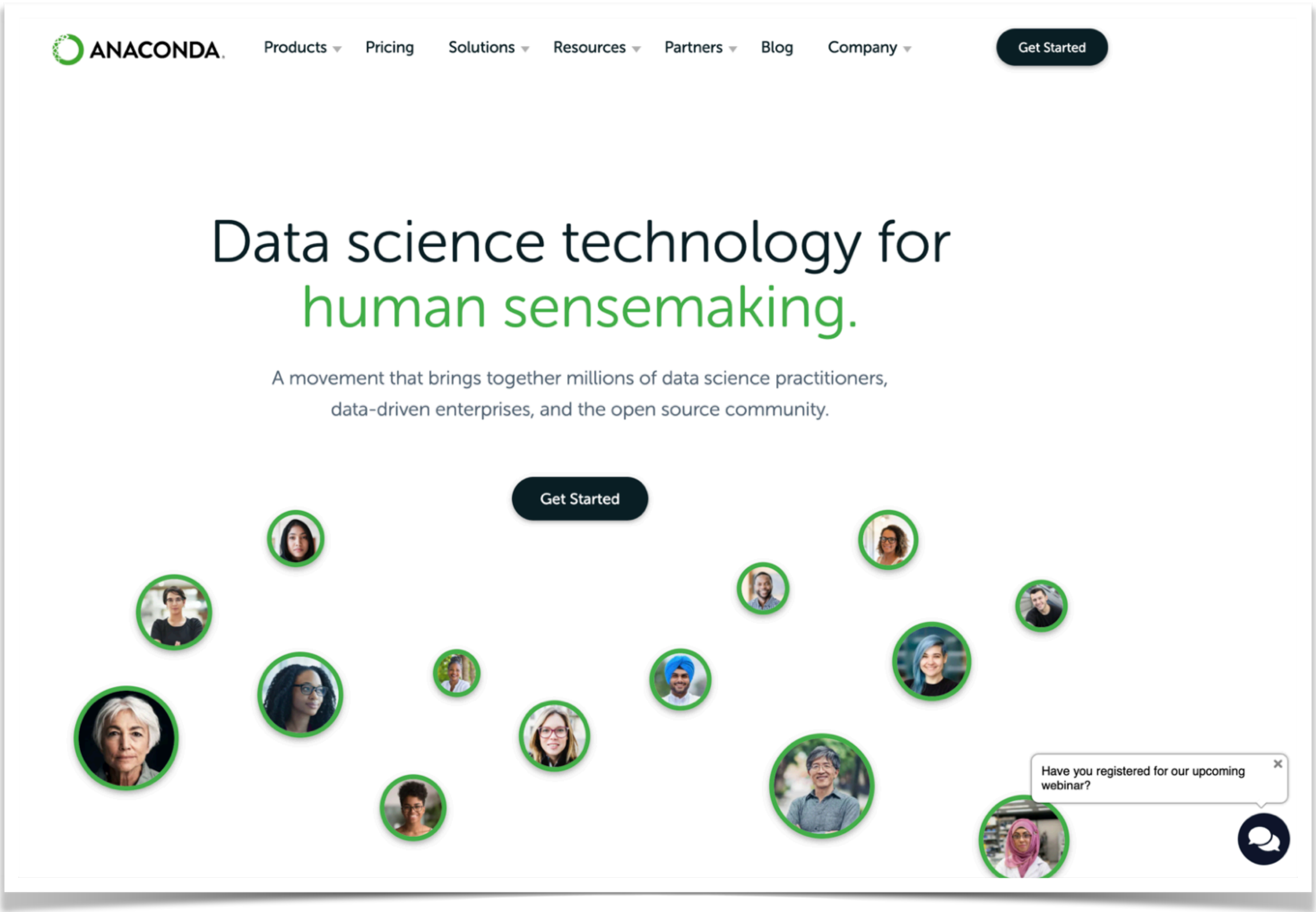
Outcome	Cour
Critique	71
Spot issues	36
Make arguments	26
Improve communication	26
See multiple perspectives	23
Create solutions	21
Consider consequences	18
Apply rules	10

Topic	Courses
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

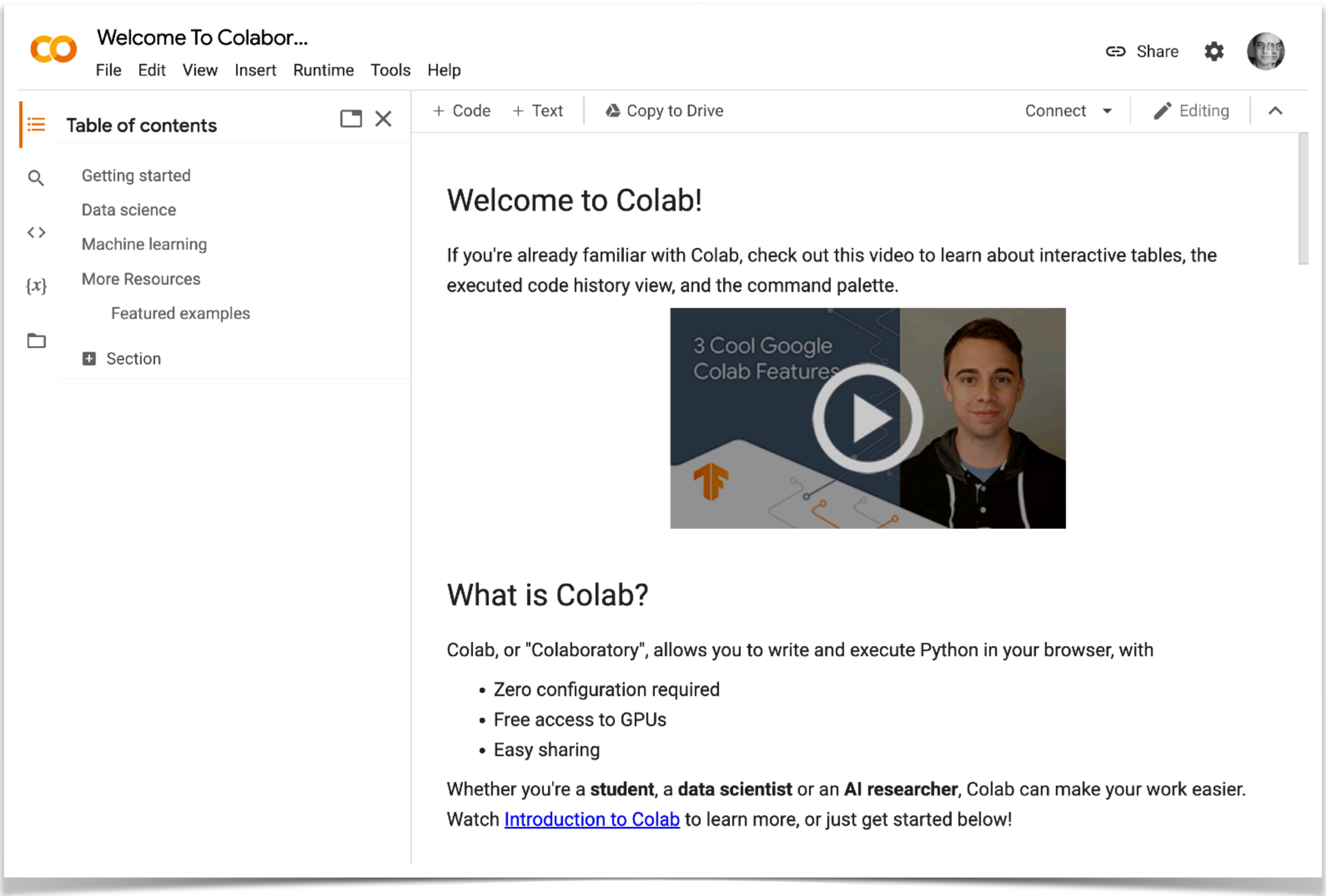
This is a hot topic!

Resources



https://www.anaconda.com/

Resources



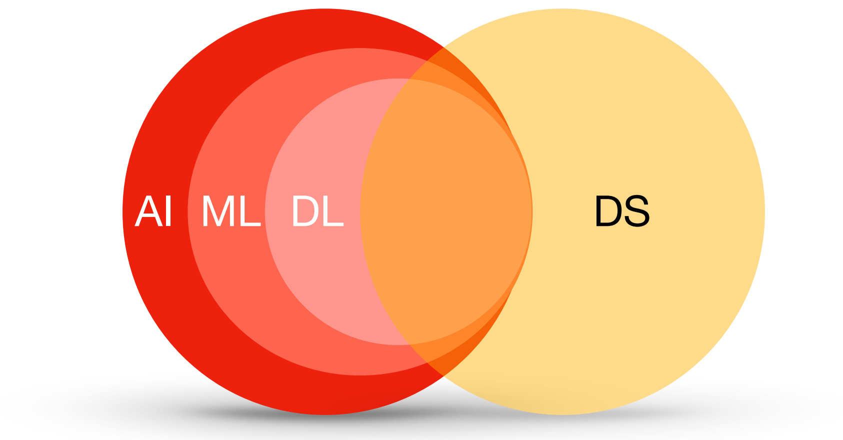
https://colab.research.google.com/

DS in Context

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.

Approach

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

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

Data Science in Context

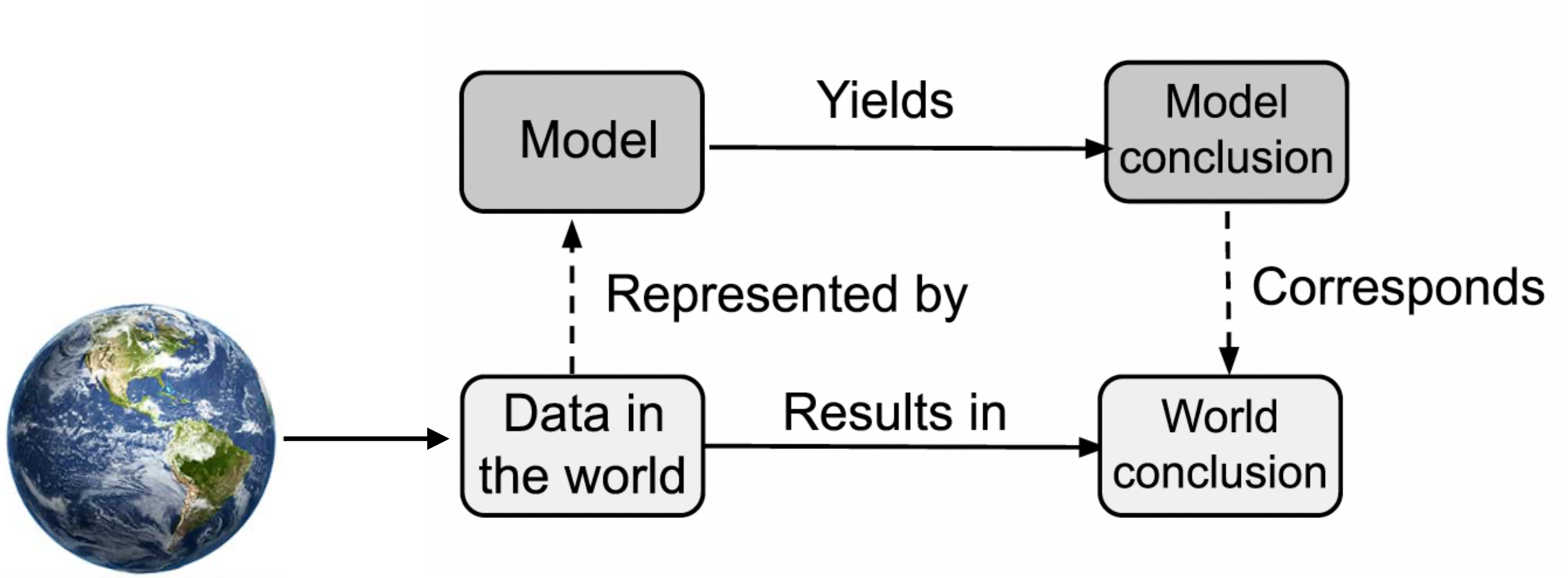
Data science is **the study of extracting value from data** – value in the form of **insights** or **conclusions**.

- A **hypothesis**, testable with more data;
- An "**aha**!" that comes from a succinct statistic or an apt visual chart; or
- A plausible **relationship** among variables of interest, uncovered by examining the data and the implications of different scenarios.
- Etc.

- **Prediction** of a consequence;
- **Recommendation** of a useful action;
- **Clustering** that groups similar elements;
- **Classification** that labels elements in groupings;
- Transformation that converts data to a more **useful form**; or
- **Optimization** that moves a system to a better state.

Data Science in Context

Insights and conclusions often arise from **models** , which are abstractions of the real world.



Models that generate these conclusions may be **clear box or black box** . A clear box model's logic is available for **inspection** by others, while an black box model's logic is not. The "opaque box" term also apply to a model whose operation is **not comprehensible**.

Data Science in Context

Data-empowered algorithms are reshaping our personal, professional, and political realities, and they are likely to have an even larger **effect** going forward.

However, as with all developing technologies, increases in impact inevitably give rise to **unanticipated consequences**.

These challenge our norms for **how we use technology in ways consistent with our values**. Many scholars, educators, and technology companies refer to these as **ethical challenges.**

Data Science in Context

Stages of <u>responsible</u> model development, debugging, understanding and deployment.

We must consider all stakeholders!

Data	Consider whether data of sufficient integrity, size, quality, and manageability exists or could be obtained.
Approach	Consider whether there is a technical approach grounded in data, such as an analysis, a model, or an interactive visualization, that can achieve the desired result.
Dependability	Does the application meet needed privacy protections? Is its security sufficient to thwart attackers who try to break it? Does it resist the abuse of malevolent users? Does it have the resilience to operate correctly in the face of unforeseen circumstances or changes to the world?
Understandability	Will the application need to detail the causal chain underlying its conclusions? Or will it make its underlying data and associated models, software, and techniques transparent and provide reproducibility?
Focus	Consider whether the application is trying to achieve well-specified objectives that align with what we truly want to happen.
Tolerance	Consider both the possible unintended side effects if the objective is not quite right and the possible damage from failing to meet objectives.
ELSI	Consider the application holistically with regard to legality, risk, and ethical considerations. Many of the topics under Dependability or Clear Objectives topics are relevant here.

Data Science in Context

Analyzing Spelling Correction

Spelling would not seem to have ethical concerns, though *Nicholas Carr* questions whether automation of mundane things is harmful to us as humans, and *Nick Romeo* questions the impact of spelling correction, per se.

- Carr, N. Is Google Making Us Stupid? The Atlantic (2008).
- Romeo, N. Is Google Making Students Stupid? The Atlantic (2014).

Data Science in Context

Analyzing Speech Recognition

There is a **privacy** issue if speech recordings and/or transcripts are transmitted or stored. Therefore, many applications do speech recognition on-device without retaining recordings.

Machine-learned speech recognition models may perform poorly for subpopulations. Supporting these subpopulations is beneficial to all, and the balance of effort to do so is a **fairness** issue that must be considered.

Data Science in Context

Music Recommendation

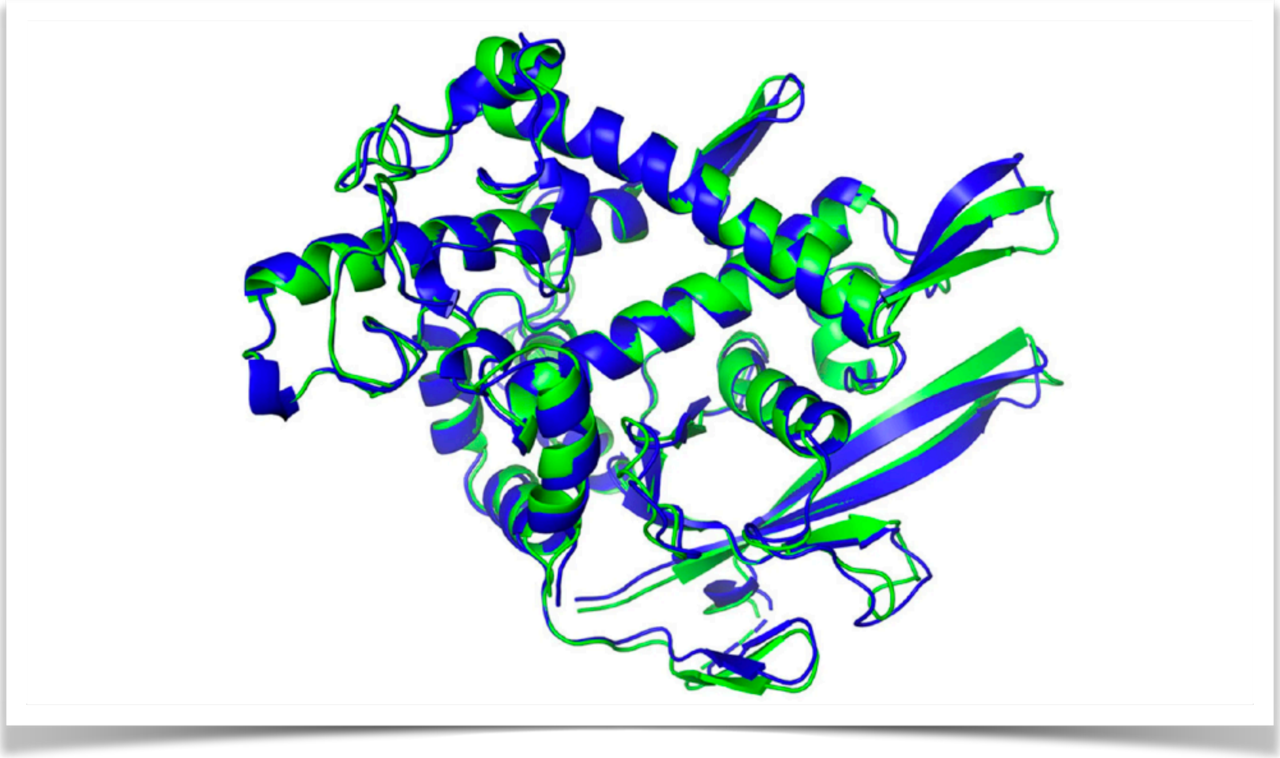
Music recommendation has few legal issues and fewer risks than other domains (although, for example, it is crucial to be careful about recommending **obscene lyrics to minors**).

However, there are many ethical issues relating to the type of recommendations made and their **impact on individual listeners, their community, and the creator/ artist** whose success may be at the mercy of these algorithms.

Data Science in Context

Analyzing Protein Folding

ELSI issues related to protein folding are minimal, though applying that knowledge (e.g., in diagnosing or treating disease) will result in many challenges.



Data Science in Context

Analyzing Healthcare Records

Health-related data is significantly **regulated**, as are study designs involving patient health records. The objectives must take account of **compliance** with these regulations.
There can be great financial and reputational risks if data is lost or misused.
Ethical issues frequently arise and are best illustrated with questions:

- Are different elements of society served equitably?
- Should a study, of potentially great value, be undertaken knowing its reproducibility might be in doubt?
- Etc.

Data Science in Context

Predicting COVID-19 Mortality

There are few, if any, legal issues.

The risks of poor forecasts are very real due to their serious impacts on health and welfare.

Data Science in Context

Intelligent transportation applications

Table 6.1 Transport & Mapping Applications of Data Science

Transport & mapping applications	Tractable data	Feasible technical approach	Depend-ability	Under-standability	Clear objectives	Toleration of failures	ELSI
Traffic speed	✓	✓	Feasible, but risks	✓	Subtle challenges	Individual but not system-wide	✓
Route finding	✓	✓	Feasible, but risks	✓	Nuances & potential externalities	No egregious errors & not system-wide	A few ELSI issues
Level-5 (fully autonomous) cars	✓	Feasibility unproven	Resilience challenge	Explanation likely needed	Difficult challenges	Great safety required	All difficult

Data Science in Context

Intelligent transportation applications

Data Science in Context

Web & Entertainment Applications

Table 6.2 Web & Entertainment Applications of Data Science

Web & entertainment applications	Tractable data	Feasible technical approach	Depend-ability	Under-standability	Clear objectives	Toleration of failures	ELSI
Identifying copyrighted videos	✓	✓ but not foolproof	Abuse	✓	✓	✓	✓
In-session video game personalization	✓	✓	Abuse	✓	Balance tricky	✓	Ethics, financial
Targeted, or personalized, ads	✓	✓	Privacy, security, abuse	✓	difficult	✓	Legal, risk, ethics
Video recommendations	✓	✓	✓	✓	ambiguity	✓	Complex
Web search	✓ but voluminous	✓ but very many techniques	Privacy, security, abuse	✓	Significant nuance	Certain failures serious	Legal, risk, ethics
News feed recommendations	Fake news	Diverse challenges	Resilience, abuse	Increasingly important	Significant nuance	Certain failures serious	Legal, risk, ethics

Data Science in Context

Web & Entertainment Applications

Data Science in Context

Medicine and Public Health Applications

Table 6.3 Medicine and Public Health Applications of Data Science

Medicine and public health applications	Tractable data	Feasible technical approach	Depend-ability	Under-standability	Clear objectives	Toleration of failures	ELSI
Mobility reporting by subregion during quarantine	✓	✓	Tricky privacy	✓	✓	✓	Perhaps, ethics
Vaccine distribution optimization - when limited supply	✓	Plausible ✓	✓	"Why" is needed	Numerous, conflicting	✓	Ethics
Identify disease outbreak from aggregated user inputs	✓	Plausible ✓	Abuse, resilience, privacy	Explanation, reproducibility	✓	✓	Perhaps, ethics
Disease diagnosis	Training data difficult to obtain	✓ for some diseases	resilience	Reproducibility, explanation, possibly causality	Agreeing on error rates	Wrong diagnoses very harmful	Legal, risk, ethics
Genome-wide association study	Difficult to obtain	Complicated by confounders, complexity	Privacy, security	Reproducibility, explanation, possibly causality	Agreeing on error rates	✓	Ethics
Understanding cause of a disease	Difficult to obtain	Complex	Privacy, security	Reproducibility, explanation, possibly causality	Agreeing on error rates	✓	Ethics

Data Science in Context

Medicine and Public Health Applications

Data Science in Context

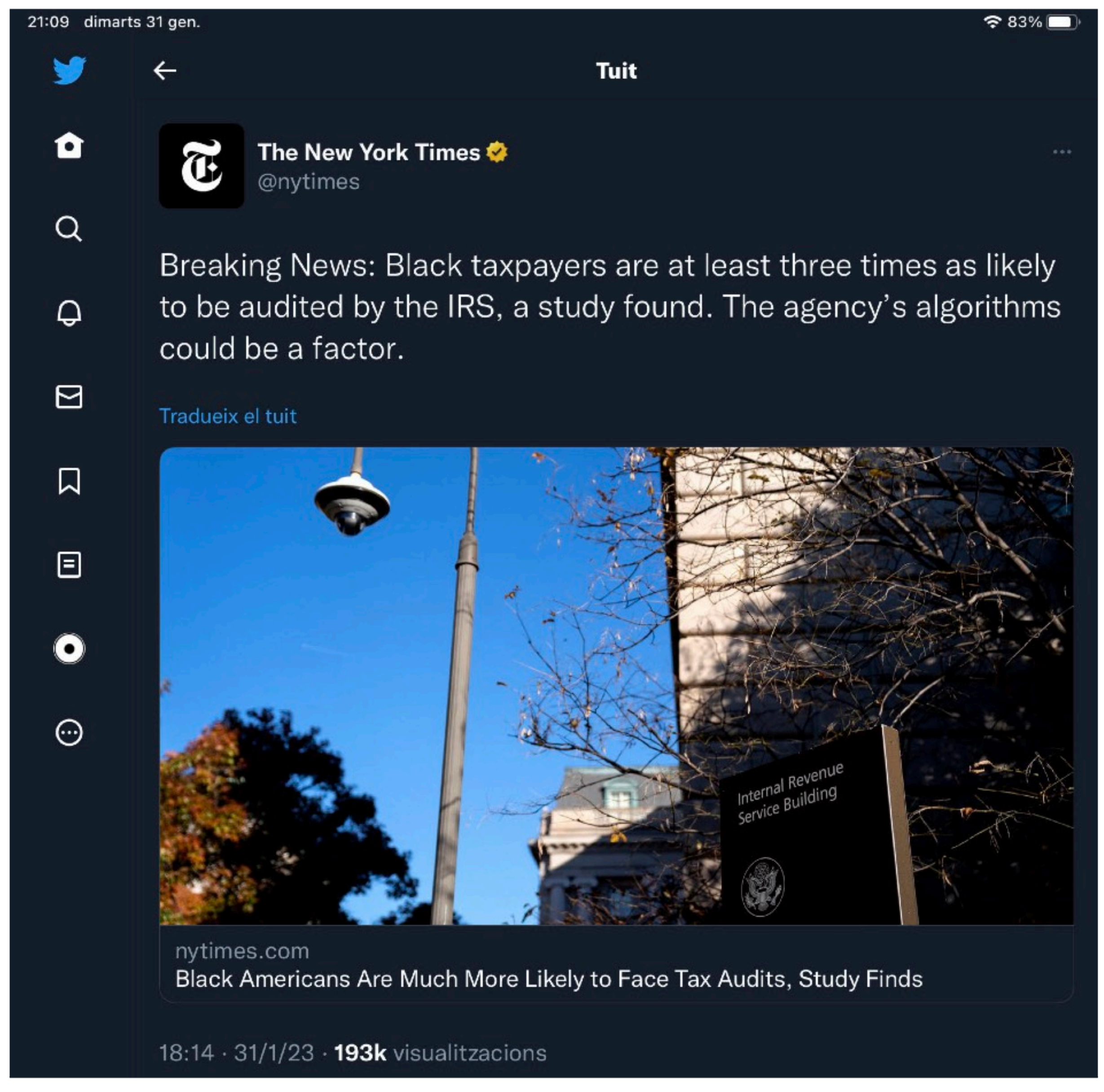
Finantial Services Applications

Table 6.5 Financial Services & Economic Applications of Data Science

Financial services and economic applications	Tractable data	Feasible technical approach	Depend-ability	Under-standability	Clear objectives	Toleration of failures	ELSI
Stock market investment selection and trading	Depends on approach	Complex but there are successes	Depends on approach	✓ opacity may be acceptable	✓	Certain failures intolerable	Legal, risk, ethics
Underwriting/pricing of property/casualty insurance	✓	✓	Privacy, security, abuse, resilience	Explanation	Competing objectives	✓	Legal, risk, ethics
"Know Your Customer" warnings for financial entities	✓	✓	Tricky Privacy, Security	Explanation	✓	Some, not all	Legal, risk, ethics
Country-wide economic prediction	Insufficient data	Feasibility unproven	Privacy, security, abuse, resilience	Explanation, reproducibility, causality	✓	Probably	Legal, risk, ethics

Data Science in Context

Finantial Services Applications



Data Science in Context

Social, Political, and Governmental Applications

Table 6.6 Government Service and Political Applications of Data Science

Government service & political applications	Tractable data	Feasible technical approach	Depend-ability	Under-standability	Clear objectives	Toleration of failures	ELSI
Targeting in political campaigns	✓	✓	Privacy, security, abuse	✓	Competing objectives	✓	Legal, ethics
Detect maintenance needs	Insufficient sensor coverage	✓	Security, resilience	✓	Complex due to prioritization	Certain failures intolerable	Legal, risk, ethics
Personalized reading tutor	✓	✓	Privacy, security, abuse, resilience	Explanation	✓	✓	Legal, risk, ethics
Criminal sentencing and parole decision-making	✓ but may be hard to assemble	✓	resilience	Explanation, reproducibility	Conflicting	Individual freedom & societal welfare	Legal, risk, ethics

Data Science in Context

Social, Political, and Governmental Applications

Preliminaries

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

Ethical relativism is the theory that holds that morality is relative to the norms of one's culture. That is, whether an action is right or wrong depends on the moral norms of the society in which it is practiced.

So, let's assume a common ground (beign aware of its limitations) based on the enlightenment, a framework that tries to encompass rationality, science, humanism and progress.

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**. (...)

A reason is an explanation of a situation or an event that provides a logical basis for a conclusion, belief, or action.

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 allowed you to exist. (...)"

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

Once upon a time...

"(...) 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!

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.

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

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

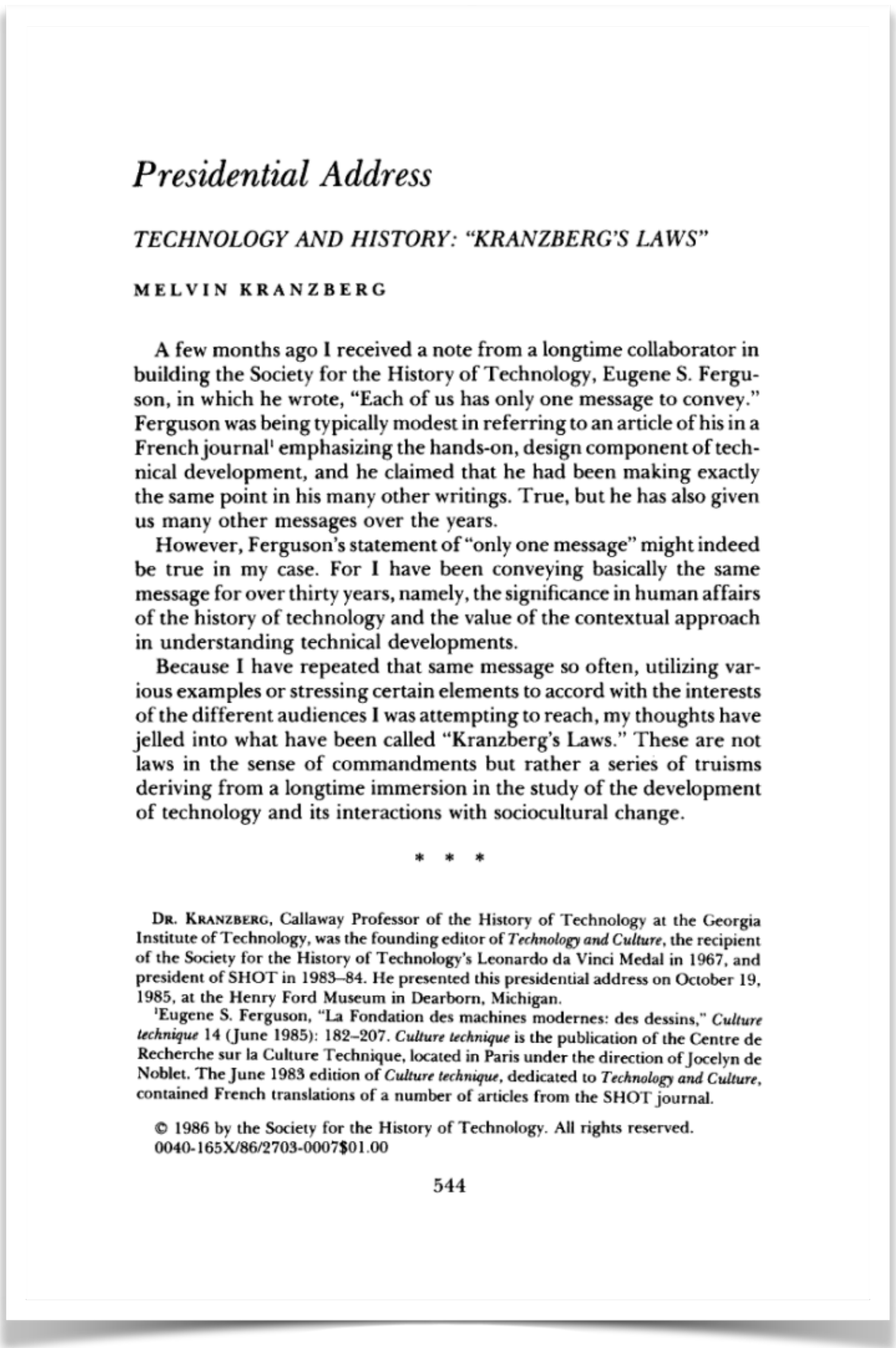
The course of **human history** can be grouped into three time periods separated by "revolutions":

- The **Cognitive Revolution** began history about [50,000, 70, 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



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 journal[1] 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.

[1]Eugene 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

Kranzberg's Six Laws of Technology

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

Kranzberg's First Law:

<u>"Technology is neither good nor bad; nor is it neutral."</u>

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.*"

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.

279

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

Original paper by Jacob Metcalf, Emmanuel Moss, and danah boyd: https://datasociety.net/wp-content/uploads/2019/09/Owning-Ethics-PDF-version-2.pdf

The **three main corporate and industry logics** in Big Tech 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**.

Original paper by Jacob Metcalf, Emmanuel Moss, and danah boyd: https://datasociety.net/wp-content/uploads/2019/09/Owning-Ethics-PDF-version-2.pdf

The role of Big Tech

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

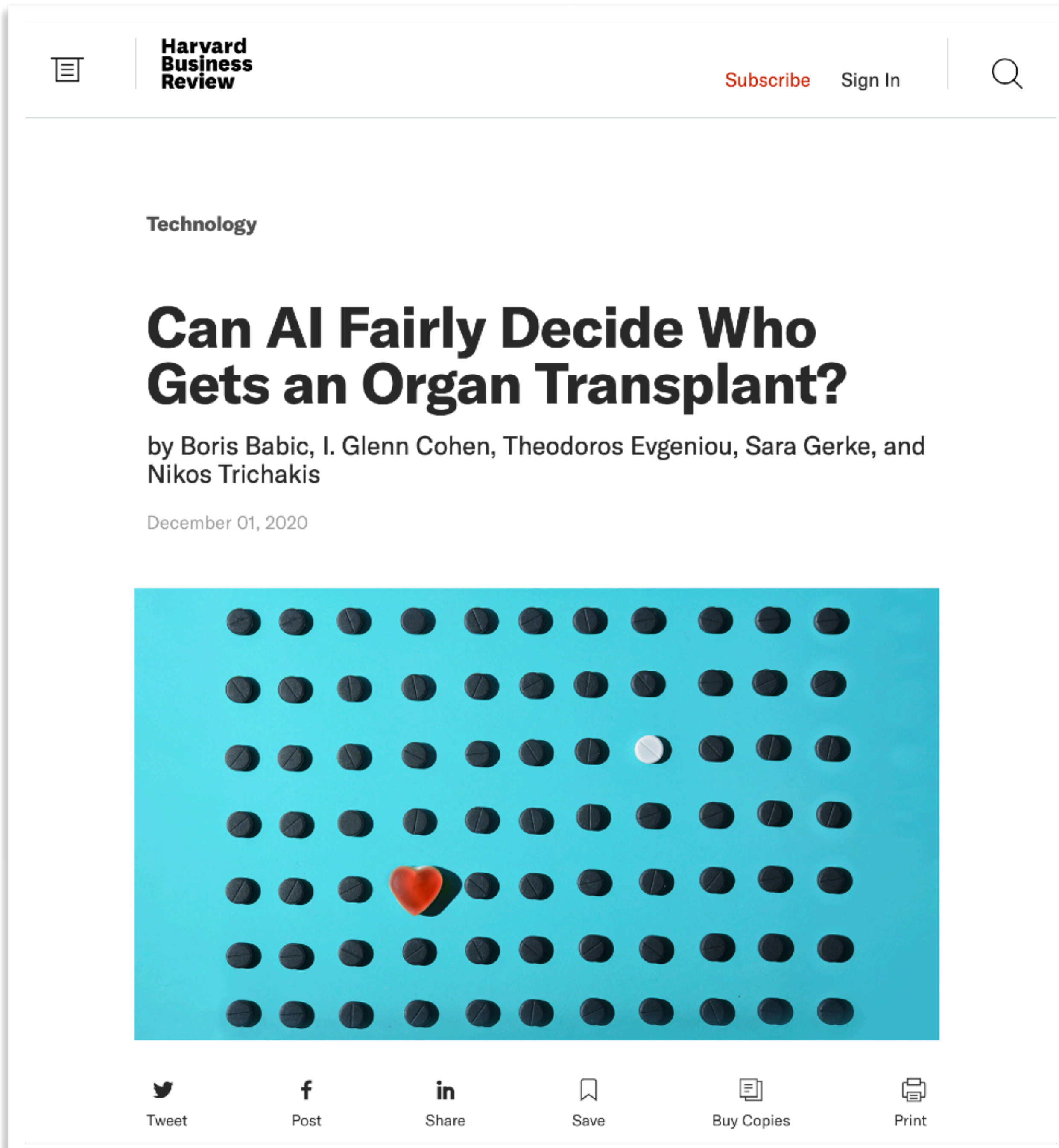
Market Fundamentalism, or market logics, refers to the idea that **companies are there to make money, and if ethics initiatives are cut into the bottom line, companies should not do it**.

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.**

Original paper by Jacob Metcalf, Emmanuel Moss, and danah boyd: https://datasociety.net/wp-content/uploads/2019/09/Owning-Ethics-PDF-version-2.pdf

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 Post Share 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

HUMAN ORGANS

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.

© 2020 The Royal Statistical Society

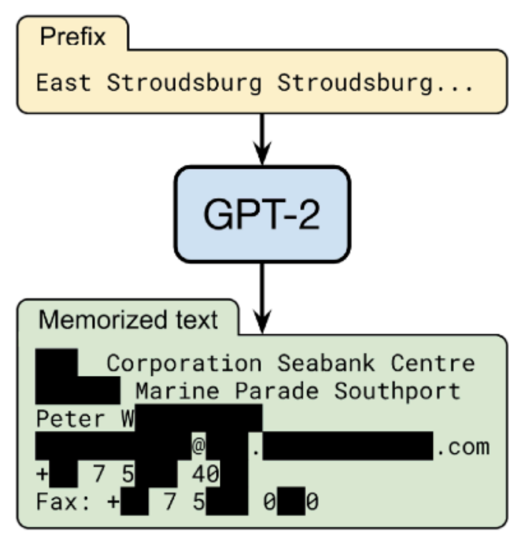
Cases

Does GPT-2 Know Your Phone Number?

Eric Wallace, Florian Tramèr, Matthew Jagielski, and Ariel Herbert-Voss
Dec 20, 2020

Most likely not.

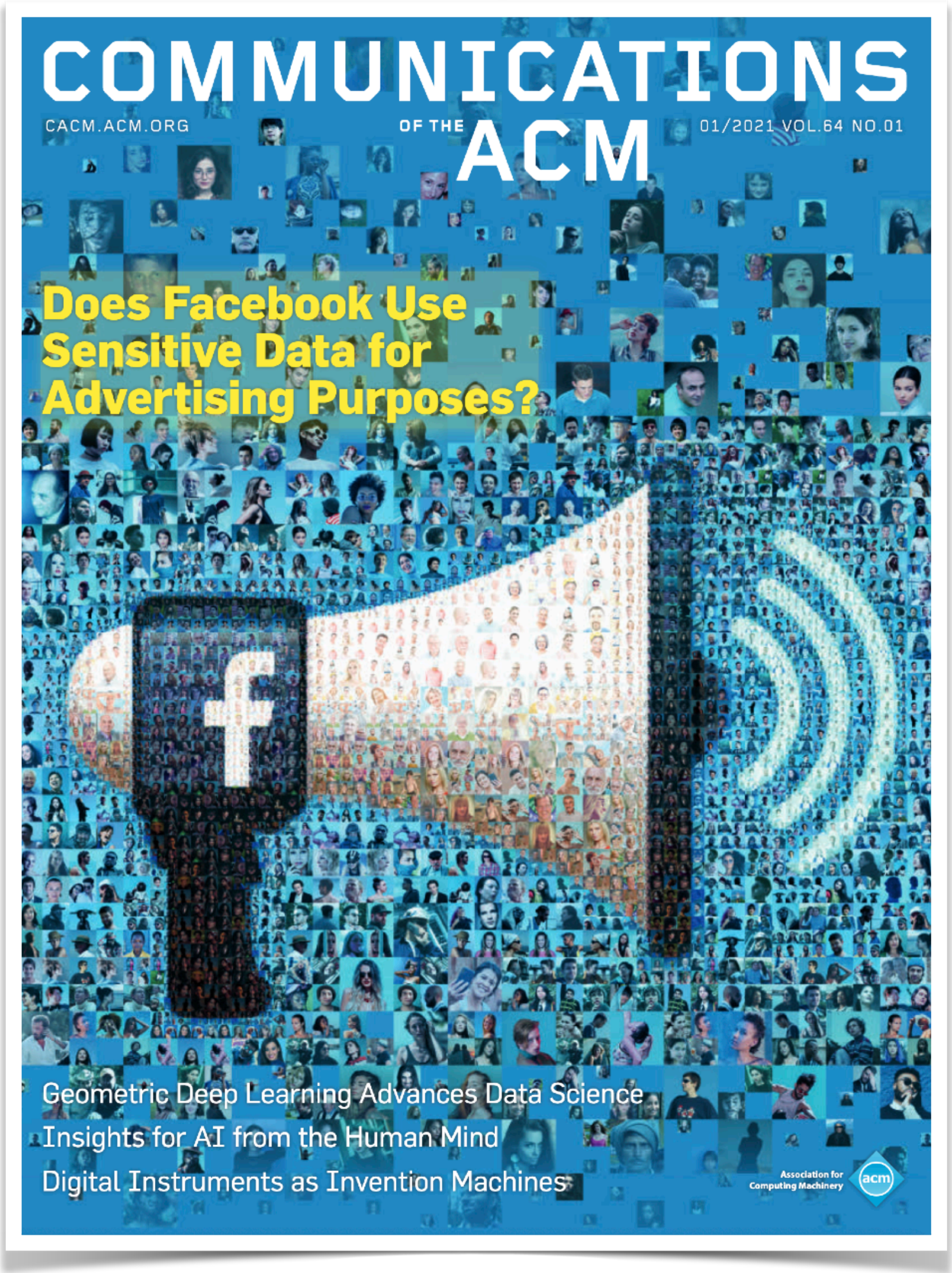
Yet, OpenAI's GPT-2 language model *does* know how to reach a certain Peter W[redacted] (name redacted for privacy). When prompted with a short snippet of Internet text, the model accurately generates Peter's contact information, including his work address, email, phone, and fax:



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, it has been 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

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

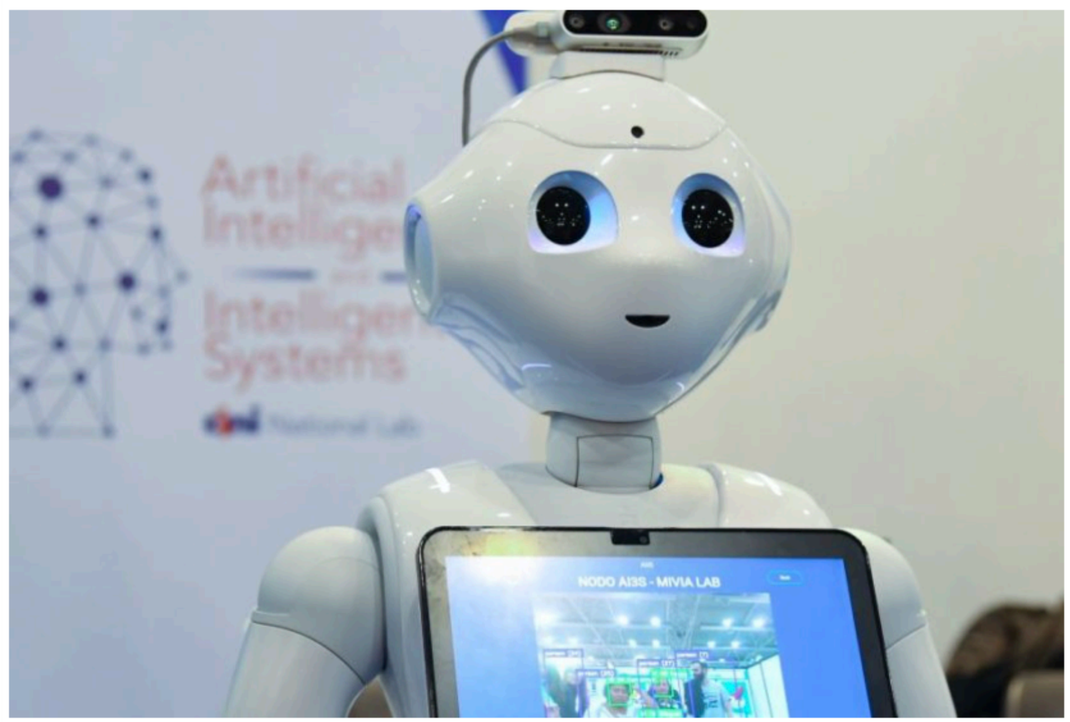
LOG IN | REGISTER

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.

Cases

POLITICO

Enter keyword

EXPLORE

SUBSCRIBE AND MORE

POLITICO PRO

REGISTER SIGN IN

HOT TOPICS | WAR IN UKRAINE | CORONAVIRUS IN EUROPE | FRENCH ELECTION 2022 | LIVING CITIES

FROM POLITICO PRO

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

Most Read Articles

1. Headscarf ban dogs Le Pen's bid for the French presidency
APRIL 18, 2022 | 4:00 AM

2. Russia's propaganda machine takes another hit
APRIL 14, 2022 | 8:59 PM

3. Le Pen vows to keep Russia close to prevent an alliance with China
APRIL 18, 2022 | 3:52 PM

4. Ukraine refugees' trauma creates 'crisis on top of a crisis' for Eastern Europe
APRIL 15, 2022 | 6:45 AM

5. Diversification from Russian gas 'possible and feasible,' says Draghi
APRIL 17, 2022 | 1:45 PM

https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/

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. Inescrutable 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?

GDPR…

- 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?

Autonomous weapons, pre-pol, AI justice,…

- Long-time/humans: What are the ethical **goals** of these technologies?

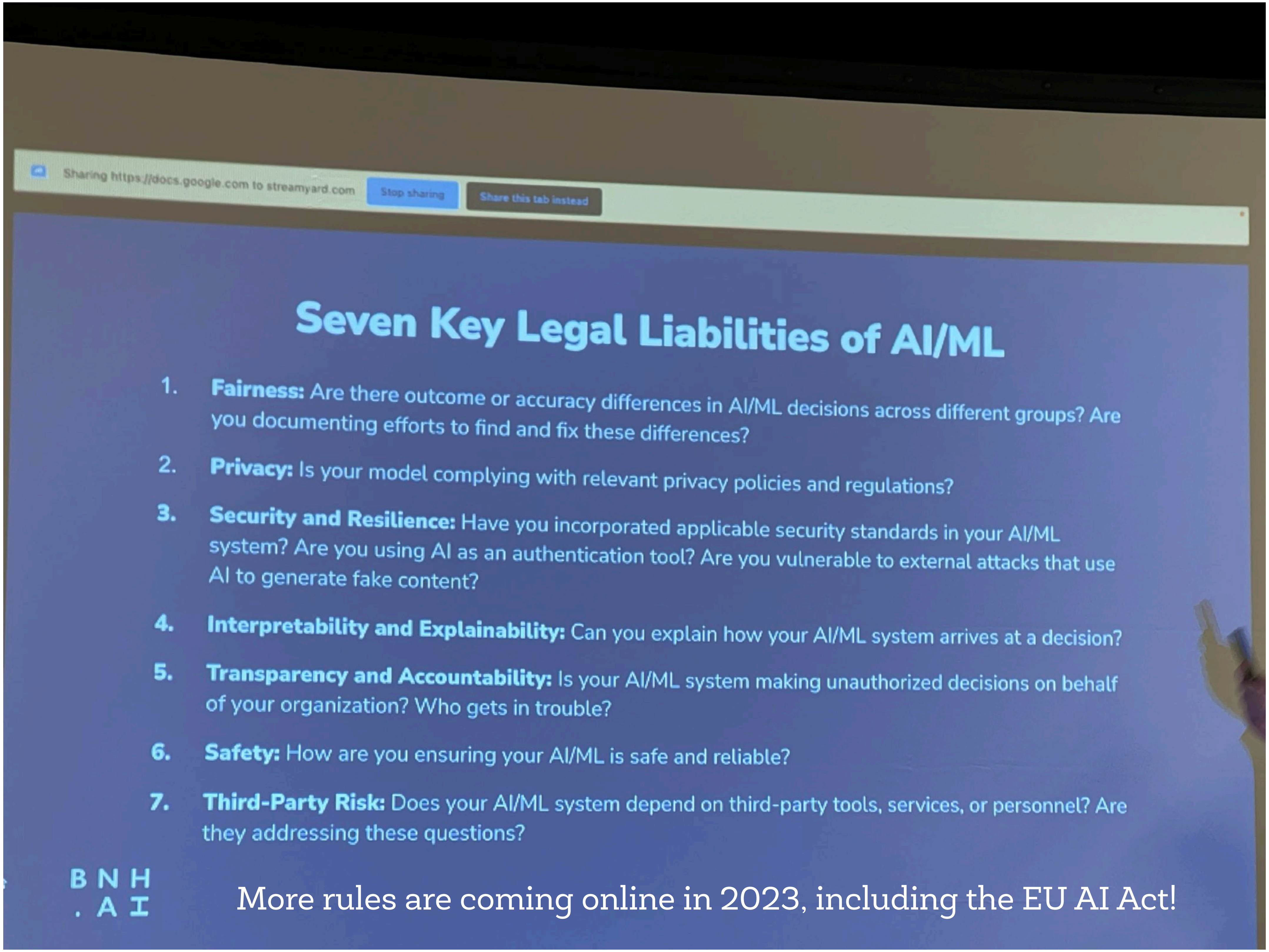
Singularity, convergence…

NLP Ethics Excuse Bingo

If I don't, someone else will	Who are you to decide?	Ethics is relative to culture	There are positive uses too
Ethics review is censorship	Science is neutral	Well, you flew to this conference	People are biased too
Negative outcomes are not predictable	Workers there are happy for $0.05	There are no alternatives	Don't slow down progress
You want to go back to candles?	Ethics review is US imperialism	The data was publicly accessible	Stop being political

CC-BY-SA

Emily M. Bender & Karën Fort 2022



More rules are coming online in 2023, including the EU AI Act!

https://twitter.com/bigdata/status/1593039585990541314/photo/4