

NLP meets the world



We have come far in NLP (and AI)

Fantastic generation quality, in different writing styles

For the first time, we have publicly usable tools

Both benchmark performance and real world evaluation by millions of users

Neural networks have done wonders for NLP

But....

“Feeding AI systems on the world’s beauty, ugliness, and cruelty, but expecting it to reflect only the beauty is a fantasy.”

Birhane and Prabhu (2021). "*Large Image Datasets: A Pyrrhic Win for Computer Vision?*",
paraphrasing Ruha Benjamin (2019)

Allocational and Representational harms

The use of AI (despite its benefits) can lead to two kinds of harms

Allocational and Representational harms

The use of AI (despite its benefits) can lead to two kinds of harms

Allocational harms arise when an automated system allocates resources (e.g., credit) or opportunities (e.g., jobs) unfairly to different social groups

- College acceptance
- Bank loan applications
- Recidivism prediction and parole

Allocational and Representational harms

The use of AI (despite its benefits) can lead to two kinds of harms

Allocational harms arise when an automated system allocates resources (e.g., credit) or opportunities (e.g., jobs) unfairly to different social groups

- College acceptance
- Bank loan applications
- Recidivism prediction and parole

Representational harms represent some social groups in a less favorable light than others, demeans them, or fails to recognize their existence altogether

- More subtle. How data is represented which leads to negative stereotypes / bias
- ... but knowledge representation is a big part of AI

Allocational and Representational harms

The use of AI (despite its benefits) can lead to two kinds of harms

Allocational harms arise when an automated system allocates resources (e.g., credit) or opportunities (e.g., jobs) unfairly to different social groups

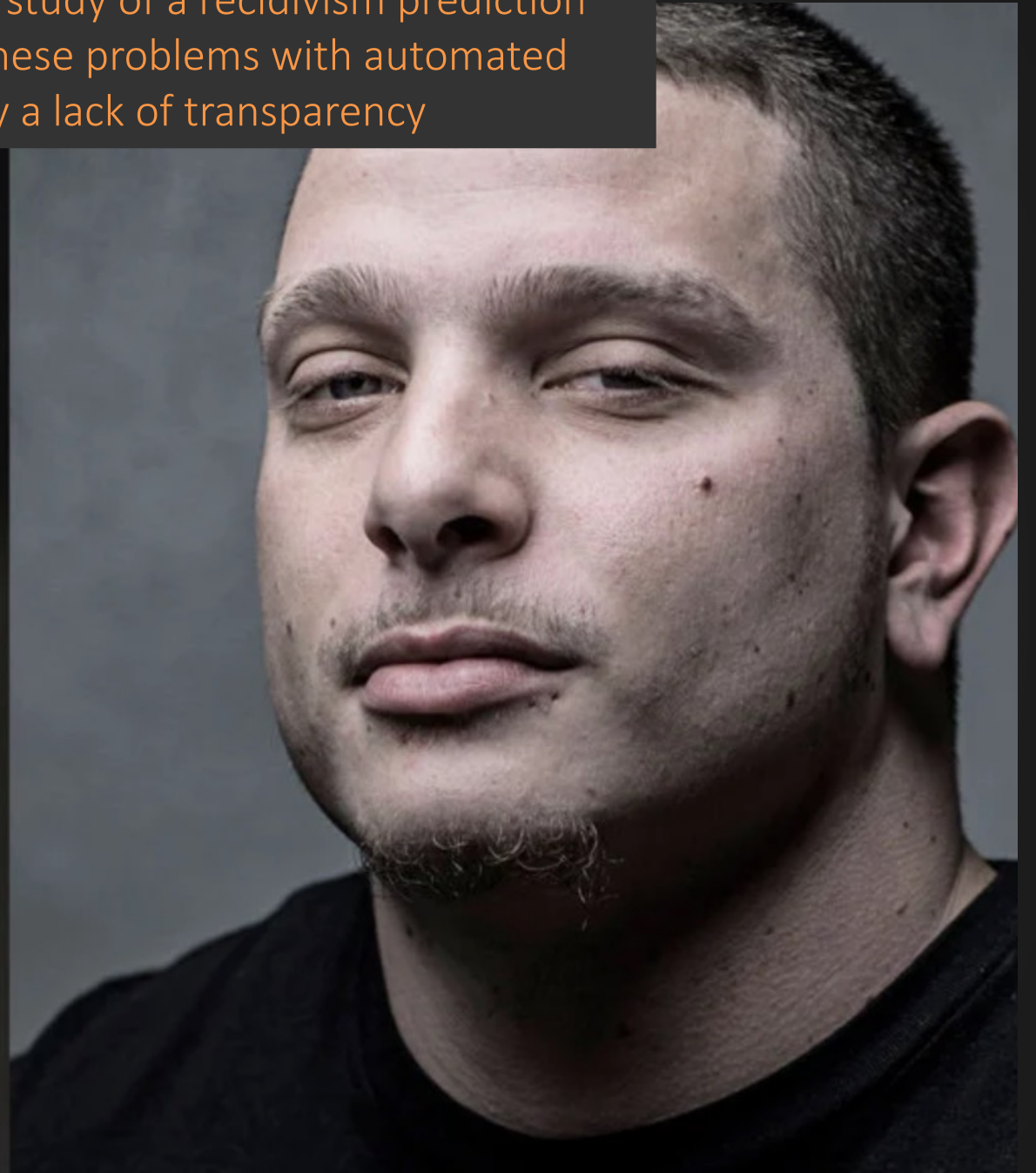
- College acceptance
- Bank loan applications
- Recidivism prediction and parole

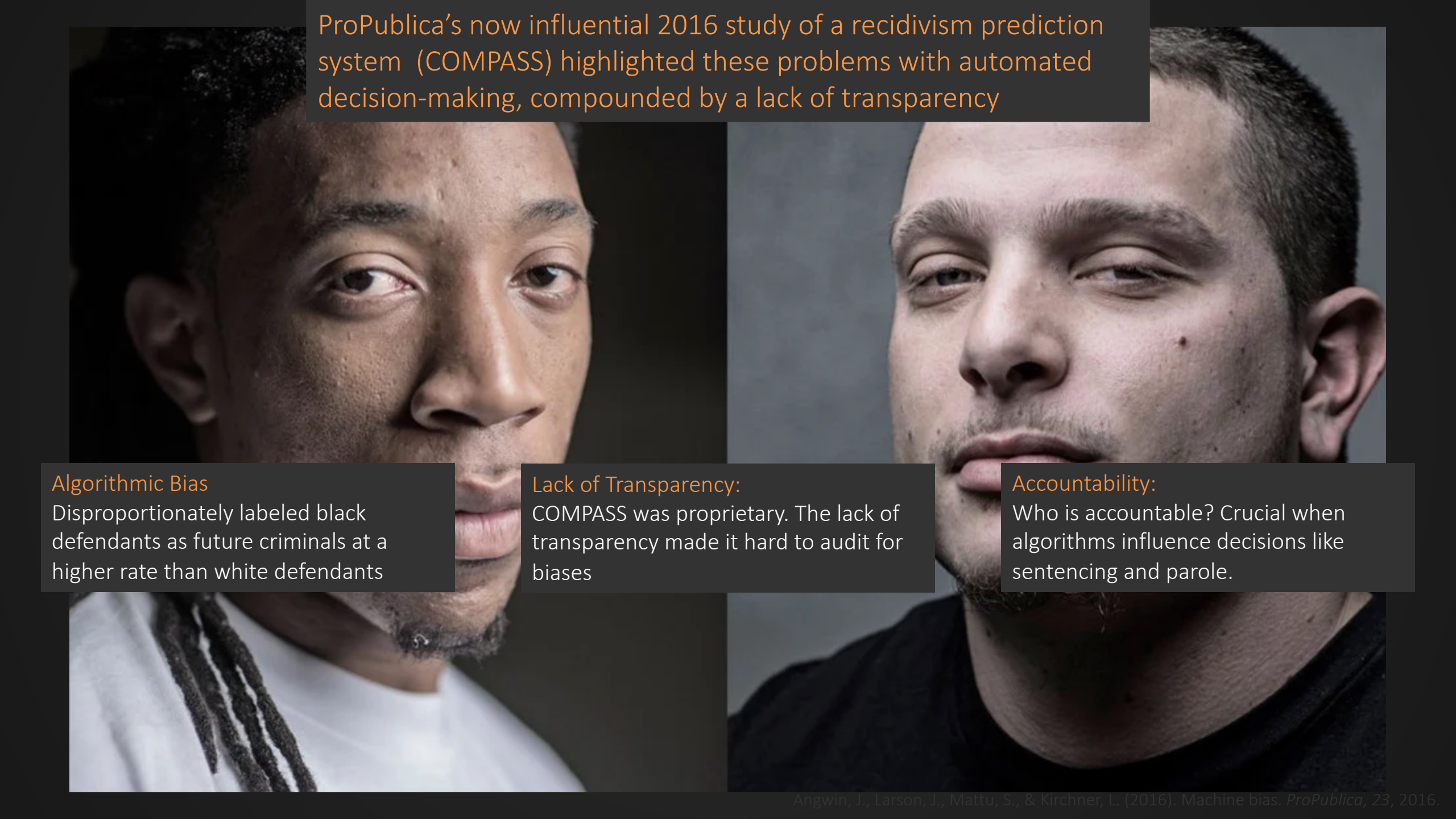
Representational harms represent some social groups in a less favorable light than others, demeans them, or fails to recognize their existence altogether

- More subtle. How data is represented which leads to negative stereotypes / bias
- ... but knowledge representation is a big part of AI

Kate Crawford's keynote at NeurIPS 2017 described this distinction. Worth looking up and watching

ProPublica's now influential 2016 study of a recidivism prediction system (COMPASS) highlighted these problems with automated decision-making, compounded by a lack of transparency





ProPublica's now influential 2016 study of a recidivism prediction system (COMPASS) highlighted these problems with automated decision-making, compounded by a lack of transparency

Algorithmic Bias


Disproportionately labeled black defendants as future criminals at a higher rate than white defendants

Lack of Transparency:

COMPASS was proprietary. The lack of transparency made it hard to audit for biases

Accountability:

Who is accountable? Crucial when algorithms influence decisions like sentencing and parole.



ProPublica's now influential 2016 study of a recidivism prediction system (COMPASS) highlighted these problems with automated decision-making, compounded by a lack of transparency

Algorithmic Bias

Disproportionately labeled black defendants as future criminals at a higher rate than white defendants

Lack of Transparency:

COMPASS was proprietary. The lack of transparency made it hard to audit for biases

Accountability:

Who is accountable? Crucial when algorithms influence decisions like sentencing and parole.

Validation

Any data-driven that makes life decisions should be validated not only for effectiveness but also for fairness, accountability and transparency



genome.gov

National Human Genome Research Institute

National Institutes of Health

Research Funding | Research at NHGRI | Health | Education | Issues in Genetics | News

Home > About > Office of the Director > Office of Population Genomics > Staff Biographies > Ebony B. Bookman

OPG: Staff Biographies

Anastasia L. Wise, Ph.D.

Ebony B. Bookman, M.S.G.C., Ph.D.

Erin M. Ramos, Ph.D., M.P.H.

Heather Junkins, M.S.

Lucia A. Hindorff, Ph.D., M.P.H.

Rongling Li, M.D., Ph.D., M.P.H.

Ebony B. Bookman, M.S.G.C., Ph.D.



Epidemiologist

[Office of Population Genomics](#)

M.S. Howard University, 1999
Ph.D. Howard University, 2001

Ads by Google

[Ebony Bookman Truth](#)

Looking for **Ebony Bookman**? Check **Ebony Bookman's** Arrests.

www.instantcheckmate.com/

[We Found Ebony Bookman](#)

1) Get **Ebony's** Background Report 2) Contact Info & More - Try Free!

www.peoplesmart.com/

[We Found Ebony Bookman](#)

Current Address, Phone and Age. Find **Ebony bookman**, Anywhere.

www.peoplefinders.com/



genome.gov

National Human Genome Research Institute

National Institutes of Health

Research Funding | Research at NHGRI | Health | Education | Issues in Genetics | News

Home > About > Office of the Director > Office of Population Genomics > Staff Biographies > Ebony B. Bookman

OPG: Staff Biographies

Anastasia L. Wise, Ph.D.

Ebony B. Bookman, M.S.G.C., Ph.D.

Erin M. Ramos, Ph.D., M.P.H.

Heather Junkins, M.S.

Lucia A. Hindorff, Ph.D., M.P.H.

Rongling Li, M.D., Ph.D., M.P.H.

Ebony B. Bookman, M.S.G.C., Ph.D.



Epidemiologist

[Office of Population Genomics](#)

M.S. Howard University, 1999
Ph.D. Howard University, 2001

Ads by Google

Ebony Bookman Truth

Looking for **Ebony Bookman**? Check **Ebony Bookman's** Arrests.

www.instantcheckmate.com/

We Found Ebony Bookman

1) Get **Ebony's** Background Report 2) Contact Info & More - Try Free!

www.peoplesmart.com/

We Found Ebony Bookman

Current Address, Phone and Age. Find **Ebony bookman**, Anywhere.

www.peoplefinders.com/

Two key issues

AI systems are increasingly adept at performing a wide range of tasks

Should increased competence warrant increased trust in an opaque decision-making system?

We are increasingly willing to deploy and use AI systems because the potential benefits are seen as important

What about high-stakes situations?

What about risks to individuals, to society, and to the environment?

Some AI models struggle with factuality

Language models can generate factually incorrect text that looks authoritatively correct at first glance

Image generation systems can create (at best) unbelievable images, and (at worst) libelous ones

How much “information pollution” is acceptable?

For entertainment applications?

For tax preparation?

Misinformation superspreaders because of the scale and easy availability?

Increased polarization?

Perhaps tacked by deeper investigative journalism

The Michael Schumacher Situation



Image credit: Ryosuke Yagi

One of the most dominant Formula One racers ever

Severely injured after a 2013 skiing accident. Reportedly in a wheelchair, paralyzed and unable to communicate

“Exclusive interview” in a German tabloid Die Aktuelle in April 2023

The entire interview was fabricated by an AI system (Character AI)

Led to public apologies, editor-in-chief’s firing, possibly a lawsuit

A misinformation superspreader?

The internet democratized the ability to spread information

Generative AI has democratized the ability to create fluent **mis**information

Together, a potent combination!

Algorithmic discrimination and data fairness

Can algorithmic decision making amplify societal biases/stereotypes?

Especially affects criminal justice, hiring, access to education and financial services approval

Does the data contain biases? Biases in the data collection process?

Whose data? Will some groups be marginalized or left behind because they are not represented in the data?

Private language models do not even reveal what data they train on, the pre-processing they use, any filters they have in the data

Privacy and trustworthiness

Is it okay to use AI systems for personal data? What about private data (e.g. medical, proprietary, etc.)?

- Would you trust a purely AI doctor or a therapist?

Can an AI model accidentally leak my private data by being trained to mimic it?

- Would you be okay if the next generation of LLMs were trained on your private data that you shared online?
- What if it produced your private data when it generated text?

Can an AI system provide sources for its claims? Explain its reasoning?

Ownership and liability

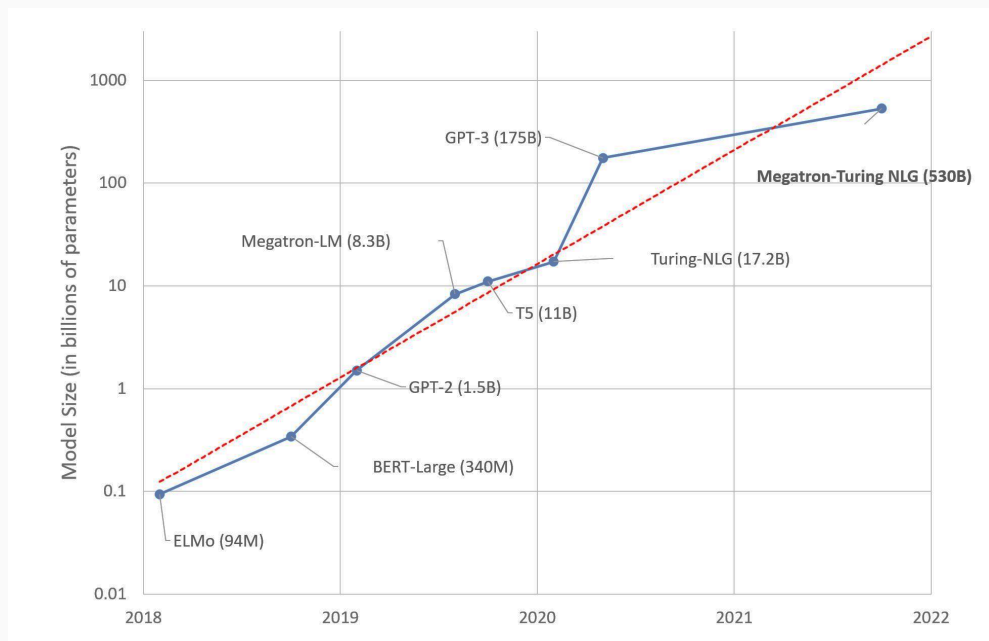
Who is the author of what an AI system generates?

Who takes ownership of the content? Who takes liability for its mistakes?

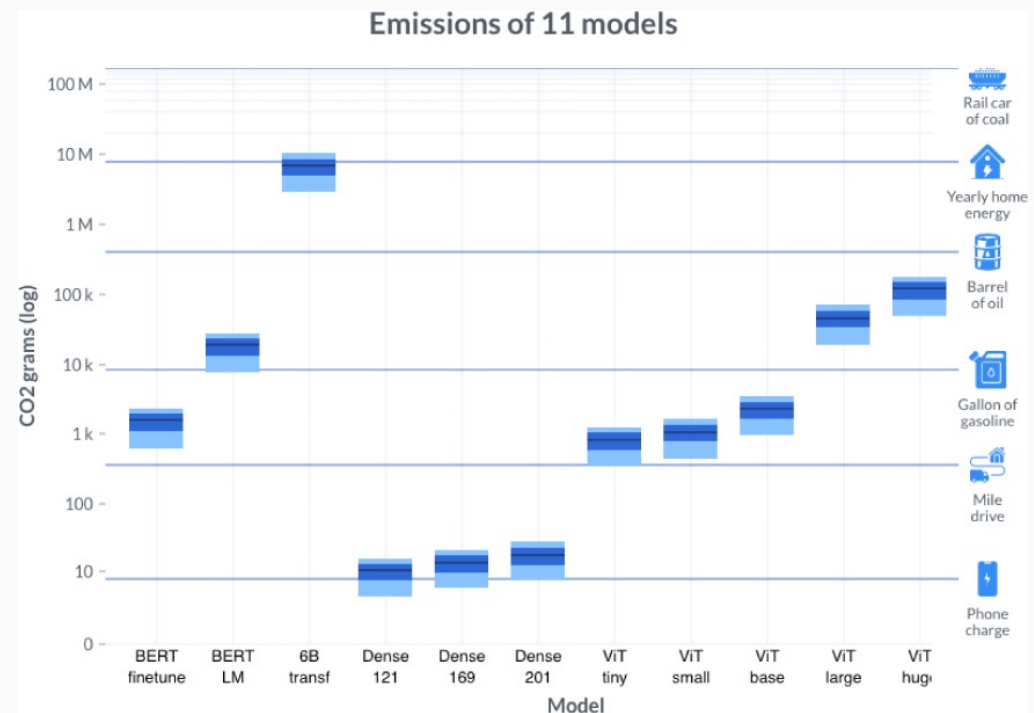
Do organizations that build and deploy AI systems bear the social costs of any harms they may cause?

Energy considerations

The largest AI systems of today require massive compute resources to train and deploy. May lead to massive energy expenditures for the compute



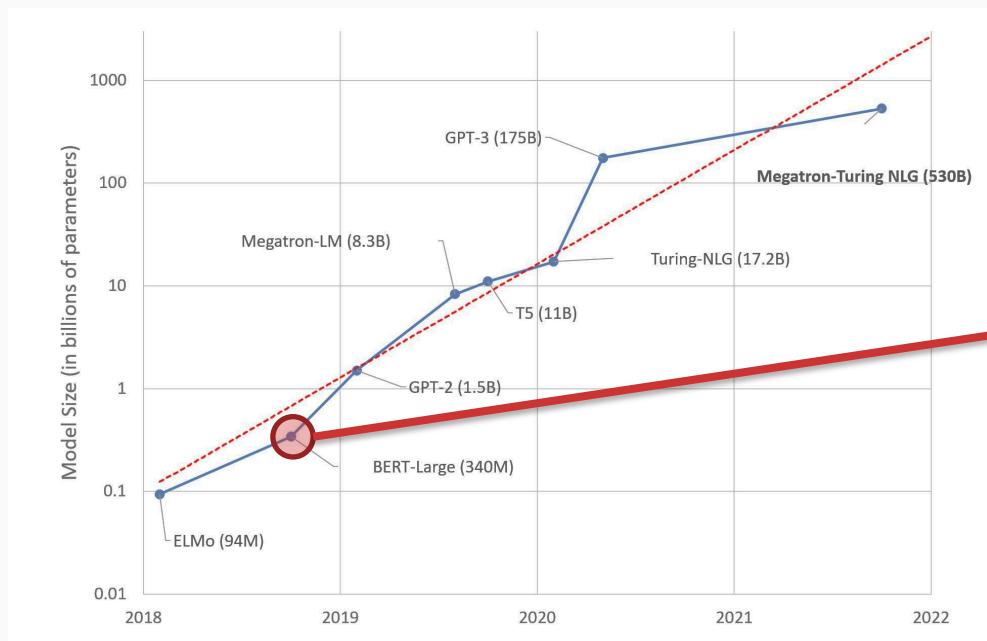
<https://huggingface.co/blog/large-language-models>



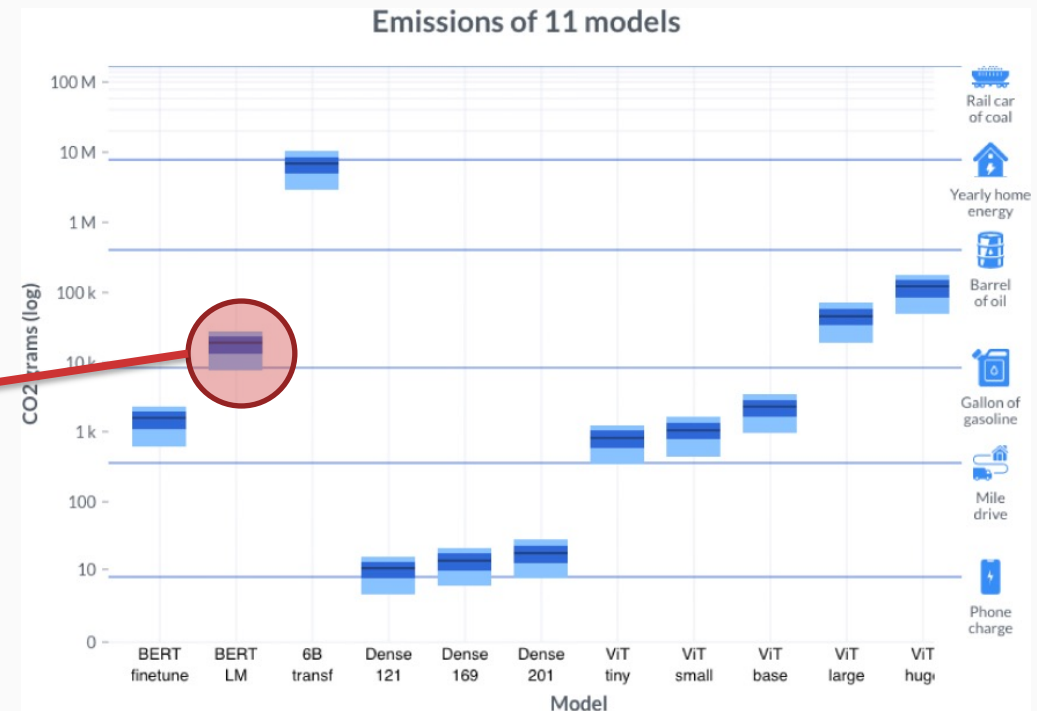
Dodge, Jesse, et al. "Measuring the carbon intensity of AI in cloud instances." *2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022.

Energy considerations

The largest AI systems of today require massive compute resources to train and deploy. May lead to massive energy expenditures for the compute



<https://huggingface.co/blog/large-language-models>



Dodge, Jesse, et al. "Measuring the carbon intensity of AI in cloud instances." *2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022.

Artificial Intelligence: A disruptor

Different for each industry and organization. What functions may be automated by data-driven compute?

Data = key asset. How to change research, education and investment priorities with this perspective?

A new “space race”. New products and AI systems being made public faster than ever

May need new governance ideas: Within organizations and beyond

AI governance

Several efforts under way across the world: EU AI law, OSTP's Blueprint for an AI Bill of Rights, NIST's AI Risk Management Framework, etc

General themes

- Effective, safe and secure systems
- Protections against discrimination of individuals or groups
- Transparent and accountable systems
- Support explainability and interpretability
- Maintain data privacy
- Human alternatives if needed

Standards organizations (perhaps new ones?) that manage and possibly enforce safeguards

AI governance

Several efforts under way across the world: EU AI law, OSTP's Blueprint for an AI Bill of Rights, NIST's AI Risk Management Framework, etc

General themes

- Effective, safe and secure systems
- Protections against discrimination of individuals or groups
- Transparent and accountable systems
- Support explainability and interpretability
- Maintain data privacy
- Human alternatives if needed

These guardrails are technological questions that need to be addressed by involving multiple stakeholders

Standards organizations (perhaps new ones?) that manage and possibly enforce safeguards

Some policy efforts under way



BLUEPRINT FOR AN AI BILL OF RIGHTS

MAKING AUTOMATED SYSTEMS WORK FOR
THE AMERICAN PEOPLE



▶ OSTP

A set of five principles and associated practices to help guide the design, use, and deployment of automated systems to protect the rights of the American public in the age of artificial intelligence

- Safe and Effective Systems
- Algorithmic Discrimination Protections
- Data Privacy
- Notice and Explanation
- Human Alternatives, Consideration, and Fallback

Diverse stakeholders need to be involved

Educators

- AI awareness in schools
- Retraining and upskilling to use AI and data-driven technology
- Integration of AI into workflow can be costly and time consuming

Government

- Ensure scientists have sufficient resources to perform research on large-scale models
- Support interdisciplinary socio-technical research on AI and its wider influences
- Encourage risk assessment when AI is developed and deployed
- Balance regulation with progress

AI researchers (both university and industry)

- Provide access to AI models and resources
- Transparency about AI tools, especially with regard to safety, fairness and reliability
- Engage with stakeholders more effectively

Write a one paragraph visual description of Salt Lake City without mentioning the name of the city



Nestled between towering mountain peaks, a bustling cityscape sprawls out across the valley below. Modern skyscrapers and historic buildings mingle together, creating a unique architectural blend. The streets are alive with activity as people go about their day, with public transportation options providing easy access to all parts of the city. In the distance, the tranquil waters of a vast saltwater lake glisten under the sun, adding to the city's natural beauty. All around, the breathtaking scenery of the surrounding mountains looms, inviting adventurers to explore the great outdoors.

Write a one paragraph visual description of Salt Lake City without mentioning the name of the city



A [artist-name] painting. Nestled between towering mountain peaks, a bustling cityscape sprawls out across the valley below. Modern skyscrapers and historic buildings mingle together, creating a unique architectural blend. The streets are alive with activity as people go about their day, with public transportation options providing easy access to all parts of the city. In the distance, the tranquil waters of a vast saltwater lake glisten under the sun, adding to the city's natural beauty. All around, the breathtaking scenery of the surrounding mountains looms, inviting adventurers to explore the great outdoors.

Write a one paragraph visual description of Salt Lake City without mentioning the name of the city



A [artist-name] painting. Nestled between towering mountain peaks, a bustling cityscape sprawls out across the valley below. Modern skyscrapers and historic buildings mingle together, creating a unique architectural blend. The streets are alive with activity as people go about their day, with public transportation options providing easy access to all parts of the city. In the distance, the tranquil waters of a vast saltwater lake glisten under the sun, adding to the city's natural beauty. All around, the breathtaking scenery of the surrounding mountains looms, inviting adventurers to explore the great outdoors.

Stable Diffusion produces

Picasso



Van Gogh



Rothko



Salvador Dali



The Times They
Are a-Changin'



Meet the new boss
Same as the old boss

