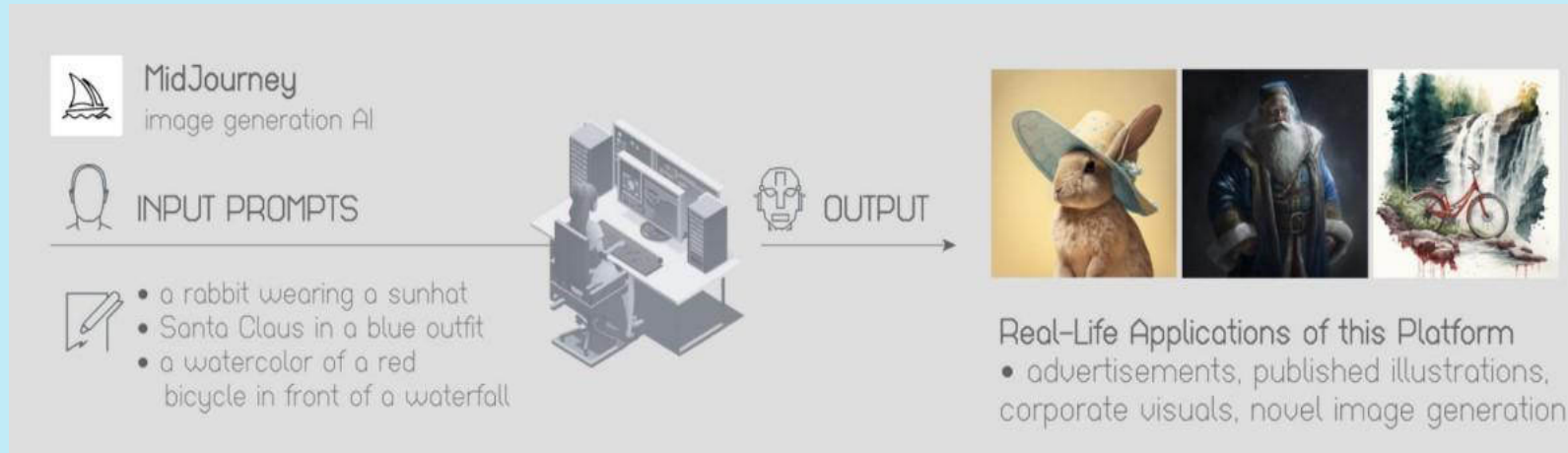


Intelligenza Artificiale: opportunità e sfide etiche

IVANA BARTOLETTI
WWW.IVANABARTOLETTI.CO.UK



Straight from the news... Generative AI (Gen-AI) is a specific type of AI that is focused on generating new content, such as text, images, music etc. using deep learning algorithms like GAN, Transformers etc.



**From Discriminative AI
Predictive models** (observe
and classify patterns in content) to...

...Generative AI (rather than
simply perceive and classify a photo of a cat,
machine learning is now able to create an
image or text description of a cat on demand)

And from...

From...

GPT 3: a large language model that works only on a single prompt.

To..

ChatGPT that can handle sequential prompts leading to a conversational style that is user friendly to humans.

Biased AI systems contributing to \$17bn gender credit gap in emerging markets: Study

Study conducted by Women's World Banking finds that fintech firms in emerging markets are missing out on an opportunity to reach 1 bn new customers.

KAIRVY GREWAL 14 April, 2021 11:01 am IST

BBC BBC

Facebook apology as AI labels black men 'primates'

It is the latest in a long-running series of errors that have raised concerns over racial bias in AI. 'Genuinely sorry'. In 2015, Google's...

Medical Algorithms Are Failing Communities Of Color

[Donna M. Christensen](#), [Jim Manley](#), [Jason Resendez](#)

AI Cybersecurity Europe Fintech

Poorly-Trained AI Algorithms Carry Bias That May Harm Financial Institutions; Expert Warns

by Tyler Smith © September 9, 2021

[Global Edition](#) [Artificial Intelligence](#)

Even innocuous-seeming data can reproduce bias in AI

Chris Hemphill, VP of applied AI and growth at SymphonyRM, says a good model performance may mask bias under the surface.

Recruiting AI systems under fire for excluding workers

AI recruiting tools focus on hiring efficiency rather than efficacy, according to an Accenture-Harvard Business School study. The U.N. believes AI is fostering human rights abuse.

PRO-VIGIL RELEASES "CONCERNING" REPORT ON AI BIAS IN VIDEO SURVEILLANCE

MORE THAN ONE-THIRD WOULD DO NOTHING ABOUT AI BIAS IN THEIR VIDEO SURVEILLANCE SYSTEMS, AS LONG AS THEY DETER CRIME

Twitter algorithmic bias bounty challenge unveils age, language and skin tone issues

The social media giant would not say if another algorithmic bias bounty challenge will be held.

• What is unique and different about AI?



AI Risks are dynamic

- Algorithms learn from new input data
- A model that was low-risk yesterday may be high-risk today, including in whether or not it is fair.

AI operates in an evolving legal landscape

Regulation around AI is evolving at the intersection between privacy, consumer, data protection, competition and human rights law.

AI systems are complex

- Fairness, for example, has different lenses, not just technical or legal;
- AI systems may be complex to interpret

Technology teams lack diversity

Lack of diversity impacts on the ability to identify potential bias at both design and implementation stage.

AI systems as socio – technical tools



Big data & automation power

Dimensions for Sustainability

1

**Individual
Dimension**

2

**Social
Dimension**

3

**Economic
Dimension**

4

**Technical
Dimension**

5

**Environment
al Dimension**

Impact of AI on the Environment

- The technology industry is estimated to be responsible for 3.5-4% of global greenhouse gas emissions, which is even higher than such sectors as air transport
- Predictions hold that AI may potentially consume 1/10 of the world's electricity by 2025

1

- For example, the language model GPT-3 OpenAI/Microsoft developed in 2020 has a capacity of 175 billion machine learning parameters and was trained on 500 billion words.
- Its estimated carbon emissions during this training were a massive 552 metric tons of carbon dioxide, a number some have compared to the greenhouse gas emissions of running 120 American cars over one year, to put it into perspective.

2

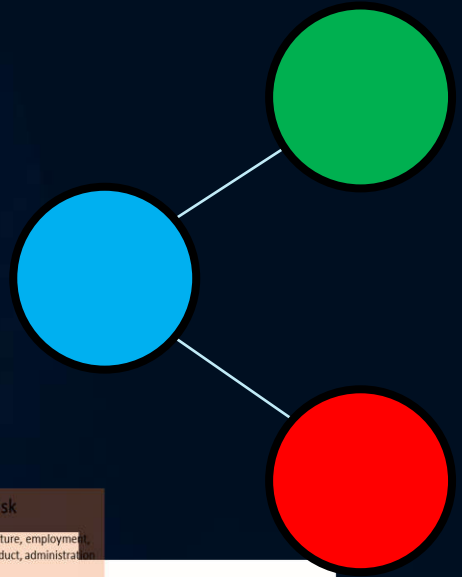
- A recent study found that, "going remote has led the global carbon footprint to grow by 34.3 million tons in greenhouse gas emissions. The scale of this increase in emissions would require a forest twice the size of Portugal to fully sequester it all
- Meanwhile, the associated water footprint would be enough to fill more than 300,000 Olympic-sized swimming pools, and the land footprint would roughly be equal in size to the city of Los Angeles.

3

Regulations around AI is coming – all around the world

2021
The EU Introduces their latest regulation around AI

The European Commission proposed a new piece of legislation that will regulate the applications of Artificial Intelligence. The legislation breaks AI into risk levels.



2019
The US Introduces regulation around algorithmic accountability

The house introduced the Algorithmic Accountability Act(AAA) bill which will empower the FTC to conduct impact assessments of algorithmic automated decision making.

Over the next couple years
Auditing and Enforcement

In the coming months, the different governmental legislation will become enforced. Companies need to start thinking how they will handle this audit process.

Bias in models can cost big dollars!

- Google will pay \$2.6 million to workers over claims its hiring and pay practices were biased against women and Asians
- IBM faces potential liability under BIPA of \$1,000 to \$5,000 per violation for each Illinois resident that had their face used for a model training.

Penalties from fines are coming

- The European Commission regulation includes fines of up to 6% of a company's annual revenues for noncompliance to the AI standard— fines that are higher than the historic penalties of up to 4% of global turnover that can be levied under the General Data Protection Regulation (GDPR).
- Local governments adding additional fines like NYC just passed a bill to fine \$1,500 per violation of companies not disclosing AI was used in hiring process.

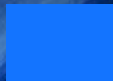
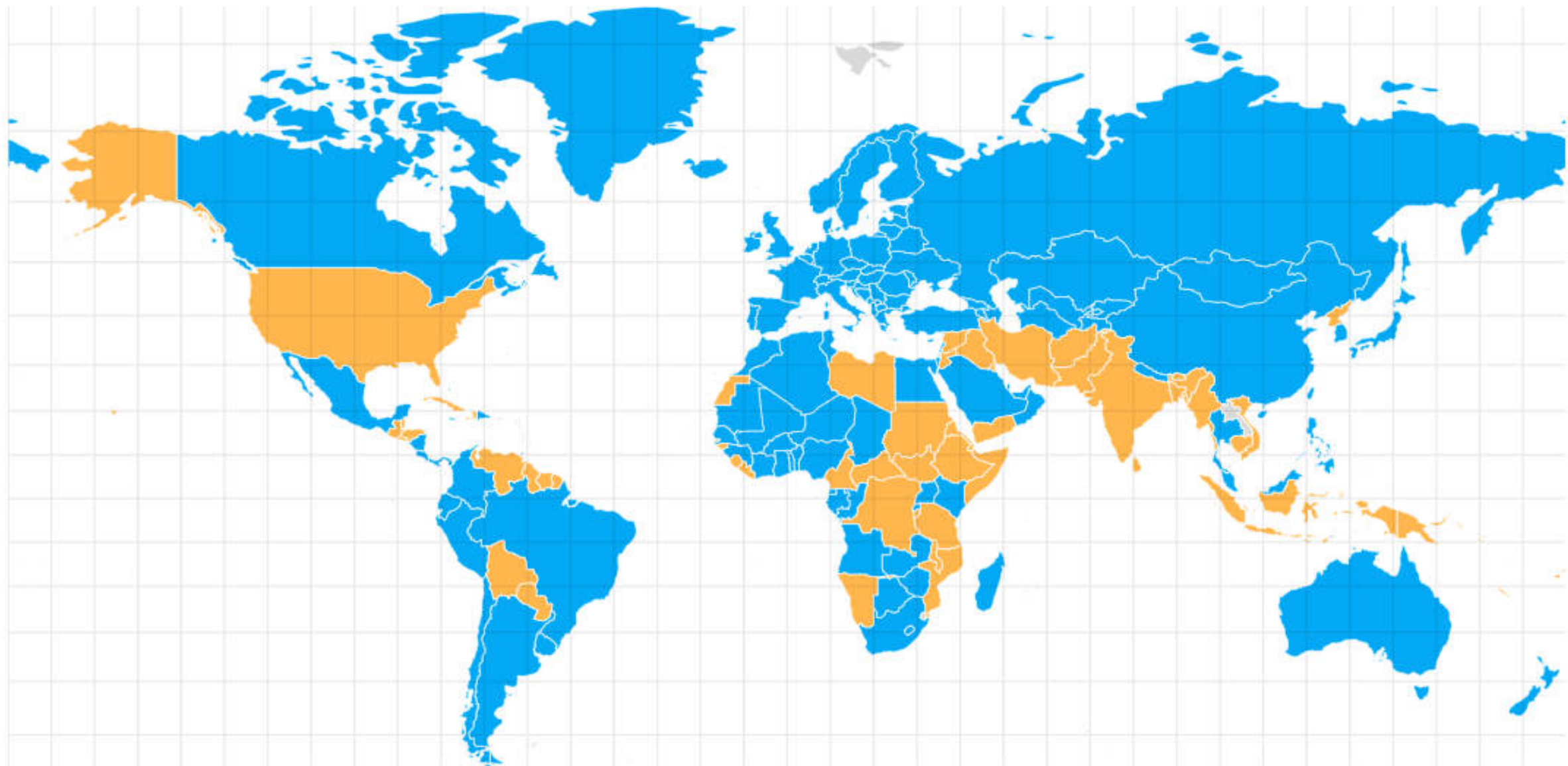
Are companies ready to comply with all the new regulations?

Global alignment on principles

Recent case law:

- *Deliveroo*
- *Ola and Uber*
- *Foodinho*
- *Schufa*
- *Syri fraud detection system*

	Proposed Taxonomy	OECD	EU	US EO 13960
Technical Design Attributes	<ul style="list-style-type: none"> • Accuracy • Reliability • Robustness • Security & Resilience 	<ul style="list-style-type: none"> • Robustness • Security 	<ul style="list-style-type: none"> • Technical robustness 	<ul style="list-style-type: none"> • Purposeful and performance-driven • Accurate, reliable, and effective • Secure and resilient
Socio-Technical Attributes	<ul style="list-style-type: none"> • Explainability • Interpretability • Privacy • Safety • Absence of Bias 	<ul style="list-style-type: none"> • Safety 	<ul style="list-style-type: none"> • Safety • Privacy • Non-discrimination 	<ul style="list-style-type: none"> • Safe • Understandable by subject matter experts, users, and others, as appropriate
Guiding Principles Contributing to Trust-worthiness	<ul style="list-style-type: none"> • Fairness • Accountability • Transparency 	<ul style="list-style-type: none"> • Traceability to human values • Transparency and responsible disclosure • Accountability • 	<ul style="list-style-type: none"> • Human agency and oversight • Data governance • Transparency • Diversity and fairness • Environmental and societal well-being • Accountability 	<ul style="list-style-type: none"> • Lawful and respectful of our Nation's values • Responsible and traceable • Regularly monitored • Transparent • Accountable



Has Data Privacy and Protection Law



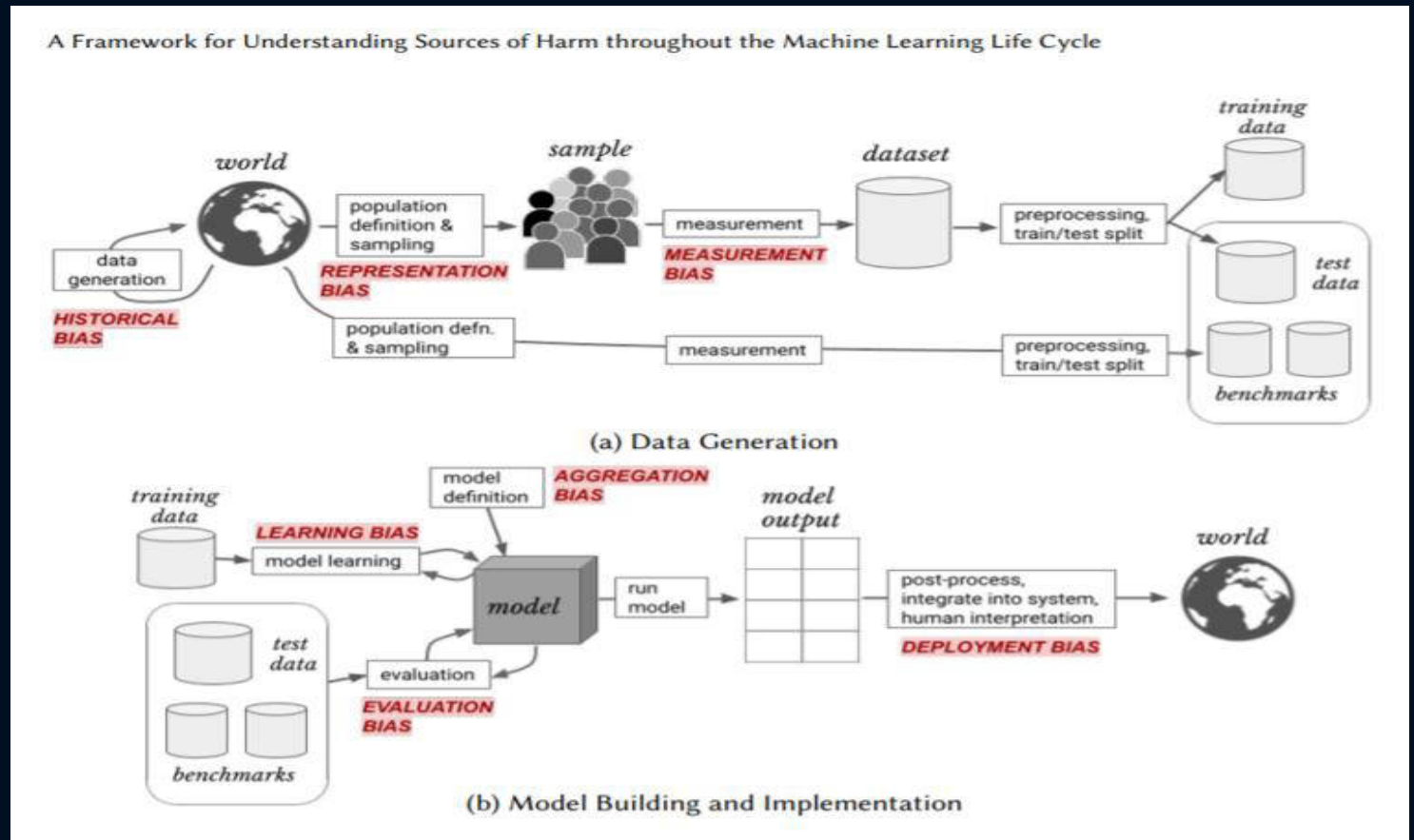
Sectoral Coverage or No Data Protection Law

Basic activity	XAI goals	Sociotechnical advice
Strategic regulation	<ul style="list-style-type: none"> • Ethical and accountability goals, both vis-à-vis employees and clients 	<ul style="list-style-type: none"> • Reach agreement on how important it is to produce fair and ethical AI solutions.
Regulation by design	<ul style="list-style-type: none"> • Pattern goals: Fairness, Privacy awareness, Causality, Transferrability, Confidence 	<ul style="list-style-type: none"> • Seek participation from operations to gain enough operational knowledge, and to bring AI design knowledge back to operations • Decide who must be involved to understand the consequences of getting it wrong • Choose a team that ensures bias awareness • Check whether the operational task structure has low degree values before redesigning the process, and check whether it will have low degree values after the redesign • Debate who to involve inside (and outside) the organization to quality control the data, to monitor how the model handles privacy, and how likely it is that the ML model can be transferred to other areas
Primary process and Operational regulation	<ul style="list-style-type: none"> • Experience goals: Accessibility, Informativeness, Interactivity, Trustworthiness 	<ul style="list-style-type: none"> • Check if there are people performing tasks in the primary processes or operational regulation • If yes, ask what kind of main structure will be used: 1. people and AI performing both primary processes and operational regulations, or 2. people performing just one of them and the AI the other? • If option 2., ensure minimizing the problems that this division of tasks produces with regards to operational explainability goals

Overview of bias: it is not just about the data!

Types of bias: historical bias, representation bias, learning bias, measurement bias, aggregation bias, evaluation bias, and deployment bias.

In a nutshell, this means that bias can emerge at any point of the AI lifecycle.



The role of diversity in AI

- Spotting bias and solutions impacting the more vulnerable
- Innovation requires deliberation
- Any algorithm built by a majority group is at risk of failing to embed perspectives of marginalised minority groups, resulting in algorithms that only work for the majority.
- Addressing diversity should be viewed as mission critical



- # Bibliography:

Wulff, K., Finnestrand, H. Creating meaningful work in the age of AI: explainable AI, explainability, and why it matters to organizational designers. *AI & Soc* (2023).
<https://doi.org/10.1007/s00146-023-01633-0>

Brevini B. Black boxes, not green: Mythologizing artificial intelligence and omitting the environment. *Big Data & Society*. July 2020. doi:10.1177/2053951720935141

Emma Strubell, Ananya Ganesh and Andrew McCallum, 'Energy and Policy Considerations for Deep Learning in Nlp', arXiv preprint arXiv:1906.02243 (2019); David Patterson et al., 'Carbon Emissions and Large Neural Network Training', arXiv preprint arXiv:2104.10350 (2021)

Patterson et al., 'Carbon Emissions and Large Neural Network Training'
Kelley Travers, MIT, ' How to reduce the environmental impact of your next virtual meeting <https://news.mit.edu/2021/how-to-reduce-environmental-impact-next-virtual-meeting-0304>