Responsible AI and Sustainability

Sasha Luccioni
Hugging Face
July 22nd, 2024

Image Credit:
Better Images of AI
(https://betterimagesofai.org/)

Hi, I’m Sasha!

I’m the Climate Lead at Hugging Face, where I work on evaluating the environmental and societal impacts of AI models and datasets.

I’m also a Board Member of **WiML** (Women in Machine Learning) and a Founding member of **Climate Change AI**

Hugging Face

WiML
Women in Machine Learning

Presentation Plan

- Introduction
- Existing Scholarship in AI Ethics & Sustainability
  - Theory
  - Practice
  - Policy
- Connecting the dots between ethics + sustainability:
  - Representativity
  - Evaluation
  - Transparency
  - Equity
- Ways Forward

Introduction

- Given their algorithmic nature, AI systems have historically been perceived as purely technical; in recent years, they are increasingly being recognized as inherently **socio-technical**
- This is due to the fact that they both **exploit** data from society and make predictions that can, in turn, **affect** its members
- AI systems are often framed within a wider context, which can vary depending on the vast array of **cultural, social,** and **environmental** contexts where they are deployed.
- This diversity can significantly impact how AI systems are **perceived** and **interacted with.**

Introduction

- For example, a system that uses AI for flood detection that was trained on data from a given region may not be suitable for use in other regions due to **differences** in the way in which cities, roads and bridges appear from space.
  - This results in **disparate impacts** depending on the context of application

Introduction

- Typically, the **ethics** and **sustainability** of AI models and systems are addressed separately
- Ethical AI principles and their operationalization consistently **overlook** systems' environmental impacts
- AI carbon footprint estimation work does not consider the **justice** or **equity** aspects of, e.g., increasingly large AI models that amplify inequality and the balance of power in the field

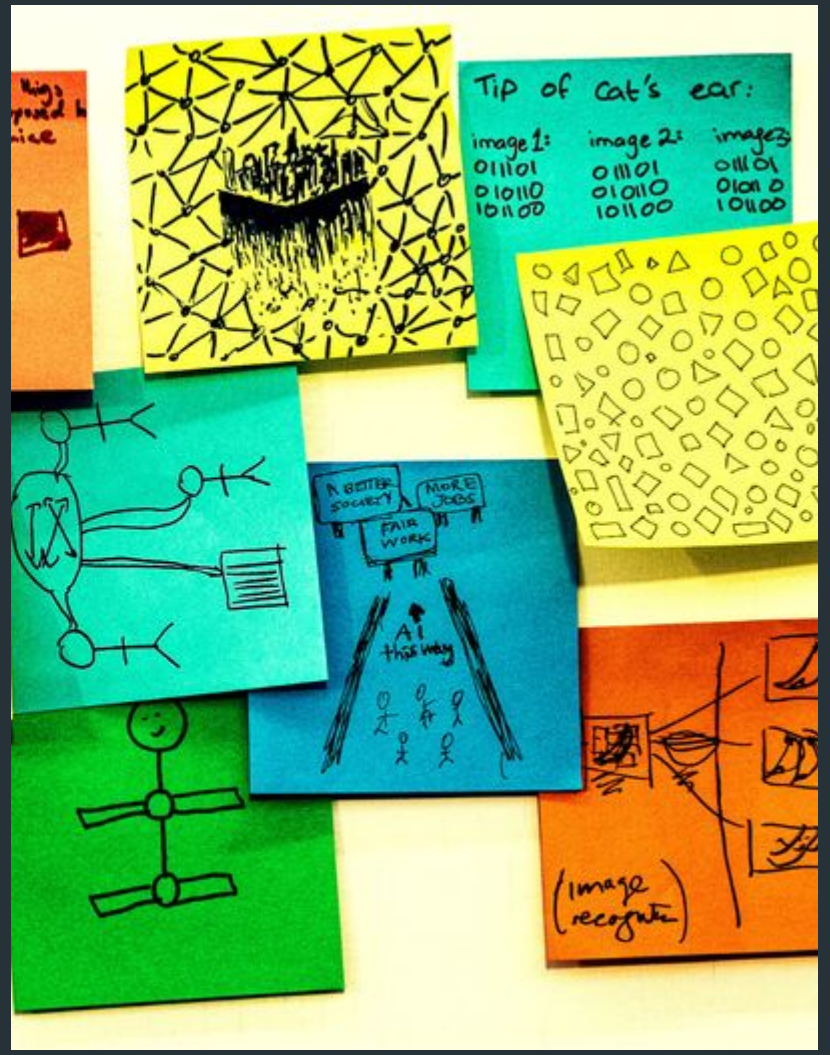
Existing Scholarship in AI Ethics & Sustainability

AI + Sustainable development

- Sustainable development: *"development that meets the needs of the present without compromising the ability of future generations to meet their own needs"* (Brundtland Report, 1987)
- Sustainable development can only be achieved when both **environmental stewardship**, **social equity** and **economic viability** coexist and are inter-connected
- These 3 pillars were the basis for the UN Social Development Goals (SDGs)
  - these have informed much of the "AI for Social Good" work in the last decade.

Principles and Frameworks

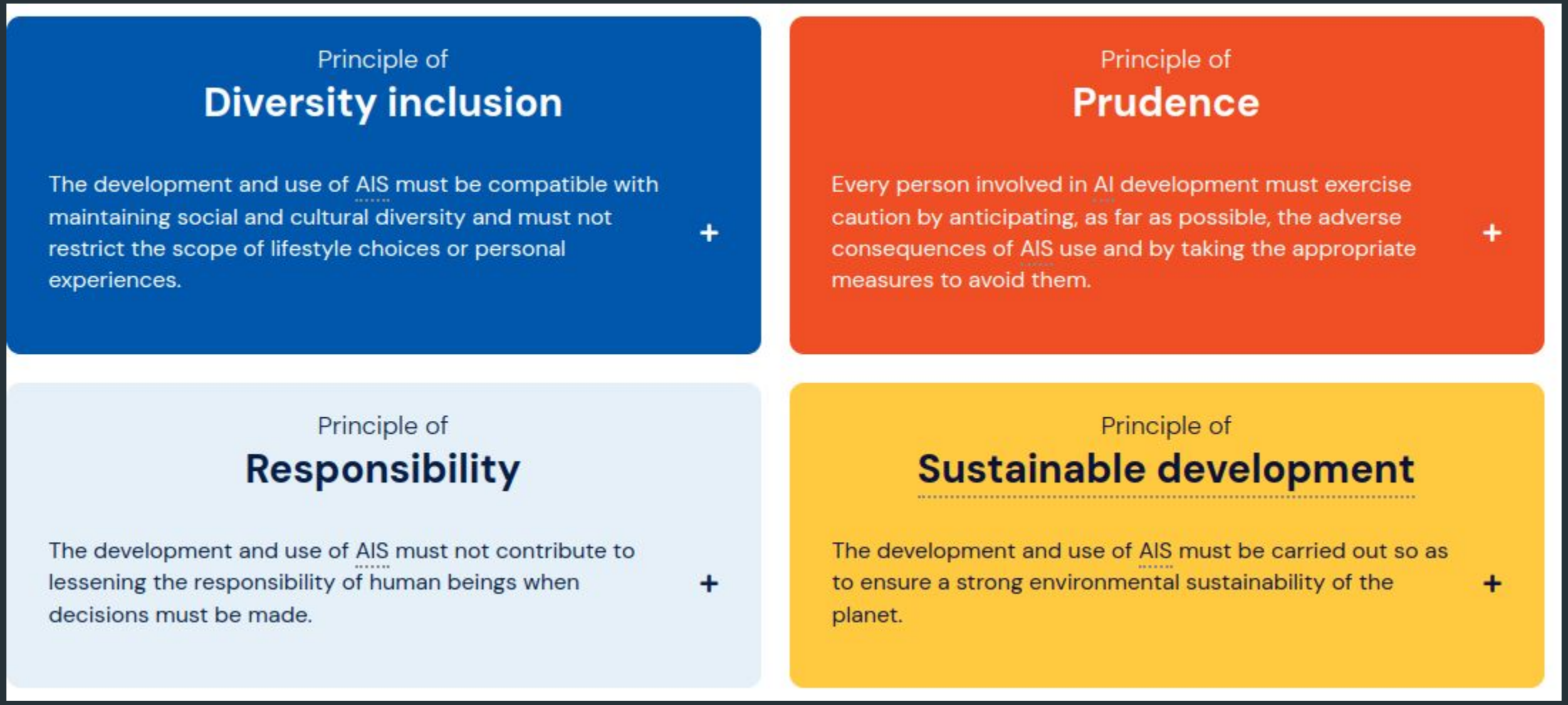
- Contemporary AI ethics has predominantly focused on informing our moral landscape and influencing future societal norms and values.
- Lacking consensus, the field often applies **Western** moral theories ranging from **utilitarianism** to egalitarianism and virtue ethics to situations involving new types of technologies:
  - autonomous vehicles
  - robots
  - chatbots



AI Guidelines

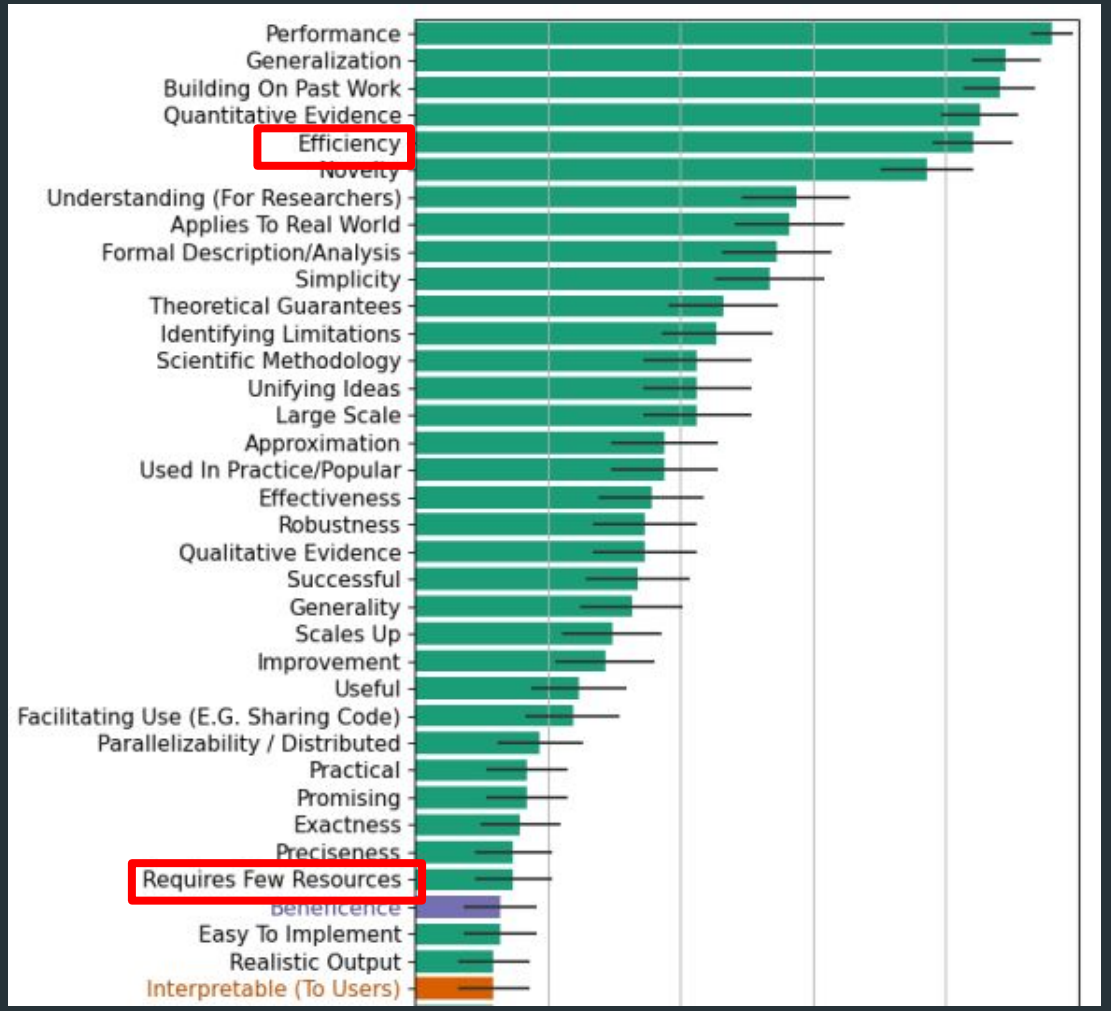
- A 2019 analysis by Jobin et al. reviewed **84 separate sets** of ethics guidelines mentioning a variety of principles, finding very **limited convergence.**
  - Of all of these guidelines, only 14 addressed sustainability, with only a single one referring to the ecological footprint of AI
- Given this divergence, AI principles and frameworks often **fall short** in addressing AI's **complex ethical issues**
- Improving things necessitates a more nuanced and **context-specific** approach to AI ethics, one that embraces the varied ethical dimensions presented by AI, including its **environmental implications.**

Montreal AI Declaration (2017)



AI & Ethics Research

- Given that AI is a **distributed field** with a multitude of **practitioners** and **organizations**, the principles that organizations strive towards can be **operationalized** differently in their day-to-day application by different members of the community.
- A 2022 study of papers submitted to ML conferences analyzed the **values** that were operationalized by their authors.
- From the 59 values they identified, **not a single one** was explicitly linked to sustainability



AI & Sustainability Research

- First research on AI + carbon footprint: Strubell et al. (2019)
- Follow-up work by other researchers has shed more light on the issue, analyzing different aspects of model architecture and its training procedure
  - This field of research has failed to take ethical considerations into account
- Even the indirect effects of using AI technologies, since they can, in fact, increase consumption, e.g. targeted advertising, AI being added to existing systems

Consumption	**CO_2e (lbs)**
Air travel, 1 person, NY↔SF	1984
Human life, avg, 1 year	11,023
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experiments	78,468
Transformer (big)	192
w/ neural arch. search	626,155

Model name	**Number of parameters**	**Power consumption**	**CO_2eq emissions**
GPT-3	175B	1,287 MWh	*502 tonnes*
Gopher	280B	*1,066 MWh*	*352 tonnes*
OPT	175B	*324 MWh*	70 tonnes
BLOOM	176B	433 MWh	25 tonnes

Life Cycle Assessment Approach

- In terms of AI & sustainability, the **life cycle assessment approach** has been proposed as a way to measure the environmental impacts of AI systems and their deployment
- While this approach does adopt a broader, 'cradle-to-the-grave' perspective, it still fails to recognize:
  - the **ethical implications** of **sustainability**
  - the contribution of large language models towards amplifying the existing **inequalities in the field of AI** due to the increased amount of computing resources and talent required to train them

(Parenthesis: Jevons Paradox)

In econom *s when*
technolog *iency with*
which a re *cost of use*
induces in *t resource*
use is incr

NeurIPS Code of Ethics (2023)

It is possible to **incentivize** ethical and sustainability considerations in AI research

Societal Impact and Potential Harmful Consequences

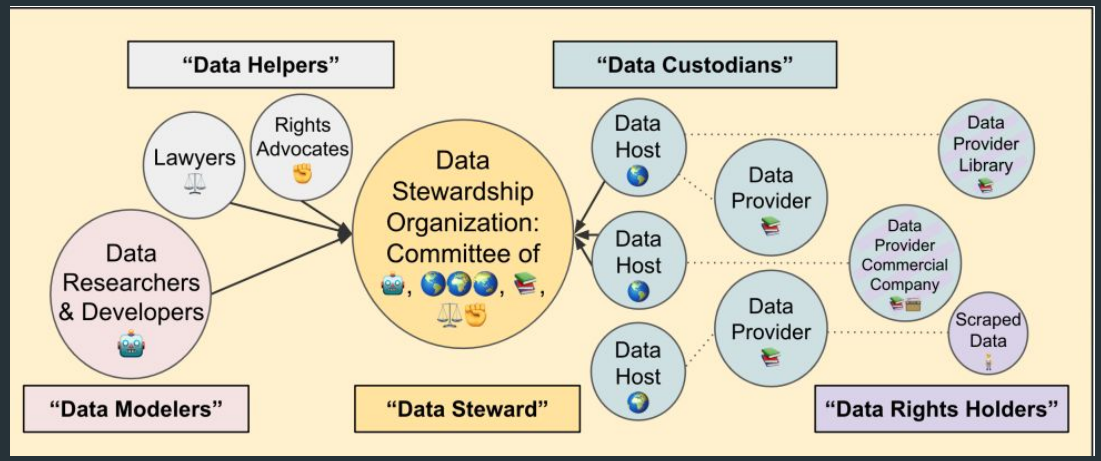
Authors should transparently communicate the known or anticipated consequences of research: for instance via the paper checklist or a separate section in a submission.

The following specific areas are of particular concern:

- *Safety:* Contributors should consider whether there are foreseeable situations in which their technology can be used to harm, injure or kill people through its direct application, side effects, or potential misuse. We do not accept research whose primary goal is to increase the lethality of weapons systems.
- *Security:* Researchers should consider whether there is a risk that applications could open security vulnerabilities or cause serious accidents when deployed in real world environments. If this is the case, they should take concrete steps to recommend or implement ways to protect against such security risks.
- *Discrimination:* Researchers should consider whether the technology they developed can be used to discriminate, exclude, or otherwise negatively impact people, including impacts on the provision of services such as healthcare, education or access to credit.
- *Surveillance:* Researchers should consult on local laws or legislation before collecting or analyzing any bulk surveillance data. Surveillance should not be used to predict protected categories, or be used in any way to endanger individual well-being.
- *Deception & Harassment:* Researchers should communicate about whether their approach could be used to facilitate deceptive interactions that would cause harm such as theft, fraud, or harassment, and whether it could be used to impersonate public figures and influence political processes, or as a tool to promote hate speech or abuse.
- *Environment:* Researchers should consider whether their research is going to negatively impact the environment by, e.g., promoting fossil fuel extraction, increasing societal consumption or producing substantial amounts of greenhouse gasses.
- *Human Rights:* We prohibit circulation of any research work that builds upon or facilitates illegal activity, and we strongly discourage any work that could be used to deny people rights to privacy, speech, health, liberty, security, legal personhood, or freedom of conscience or religion.
- *Bias and fairness:* Contributors should consider any suspected biases or limitations to the scope of performance of models or the contents of datasets and inspect these to ascertain whether they encode, contain or exacerbate bias against people of a certain gender, race, sexuality, or other protected characteristics.

AI & Ethics : Governance

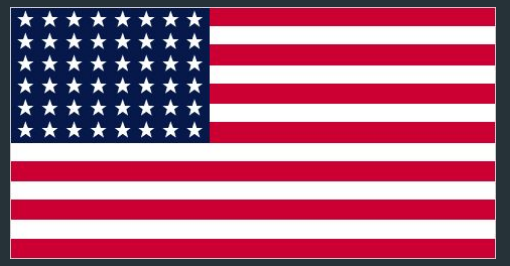
- **Governance** and **regulation** are core to both AI ethics and sustainability, and they each include both bottom-up and top-down endeavors for:
  - establishing mechanisms for **decision-making**
  - guiding the **development** and **deployment** of AI systems
  - outlining the **roles** and **responsibilities** of each party



Legislation efforts (2023/2024)

FEBRUARY 01, 2024

MARKEY, HEINRICH, ESHOO, BEYER INTRODUCE LEGISLATION TO INVESTIGATE, MEASURE ENVIRONMENTAL IMPACTS OF ARTIFICIAL INTELLIGENCE



EU AI Act: first regulation o intelligence

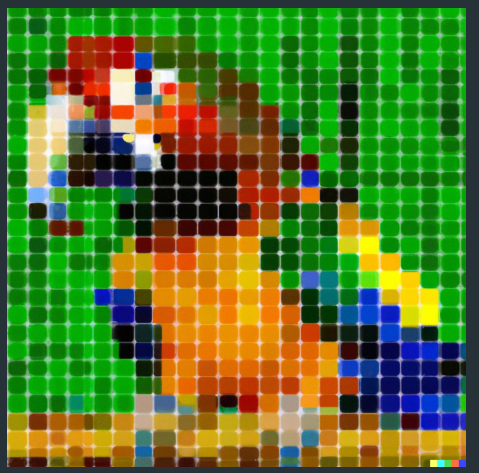
Society Updated: 19-12-2023 - 11:45
Created: 08-06-2023 - 11:40

the measurement and logging methods to enable calculations and reporting of the environmental impact of systems to comply with the obligations in this Regulation, including carbon footprint and energy efficiency, taking into account state-of-the-art methods and economies of scale.

Connecting the Dots Between AI Ethics & Sustainability

From “Stochastic Parrots” (Bender, Gebru et al., 2021)

“Is it fair … that the residents of the Maldives (likely to be underwater by 2100) or the 800,000 people in Sudan affected by drastic floods pay the environmental price of training and deploying ever larger English LMs, when similar large-scale models aren’t being produced for Dhivehi or Sudanese Arabic?”

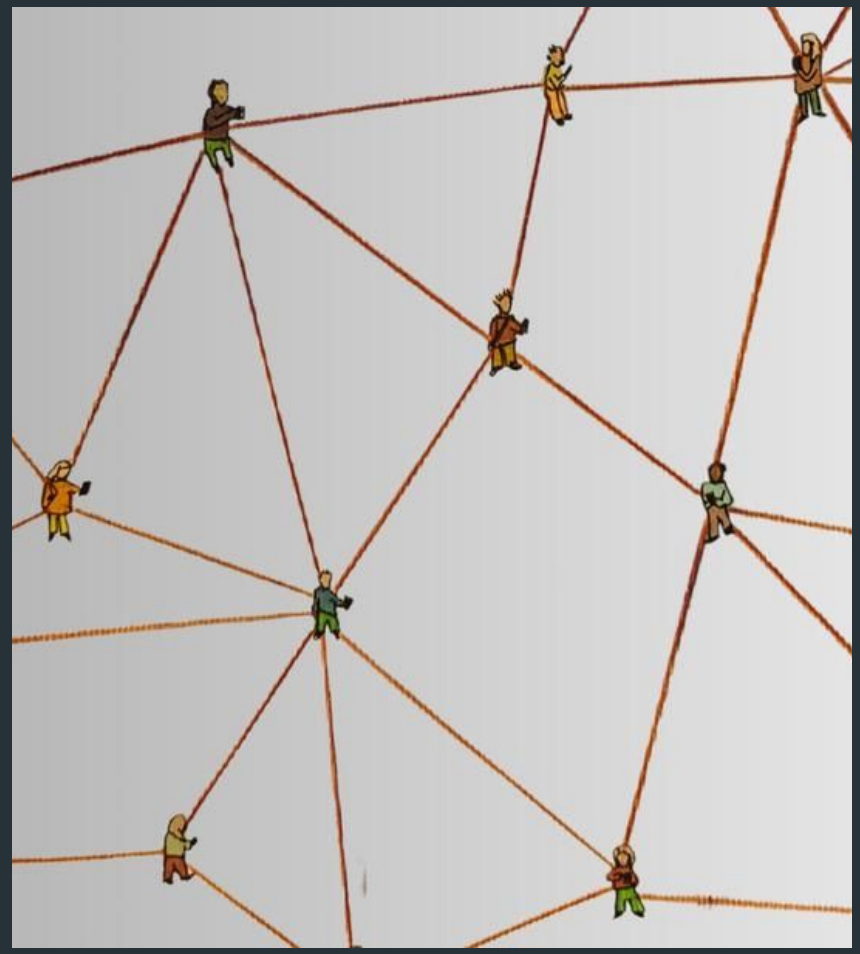


… And how would we know the extent of the broader societal and environmental impacts of LMs unless they were holistically evaluated by model creators and these results were transparently communicated to the community?

Connecting the dots between ethics + sustainability

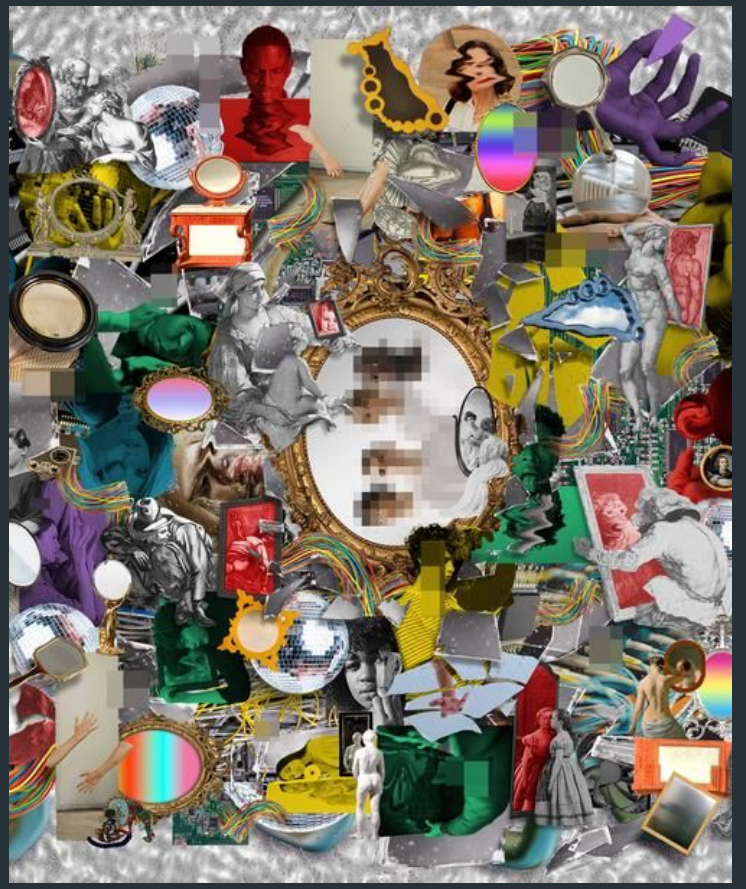
1. Representativeness
2. Evaluation
3. Transparency
4. Equity

These themes were all raised in different contexts by critical scholars such as Dobbe and Whittaker (2019), Birhane (2022) and Van Wynsberghe (2021)



Representativeness

- AI functions based on assumptions of **data representativeness**
  - that the data used for training AI models will allow them to **accurately represent** the world.
- The myth of **generalizability**, and the perception of ML models, especially LLMs, as “**general purpose technologies**'', persists.
- What the AI ethics community refers to as **bias** is often a result of flawed assumptions about **model generalizability** and data **representativity**.



Representativeness- *Bugs in the Data (2023)*

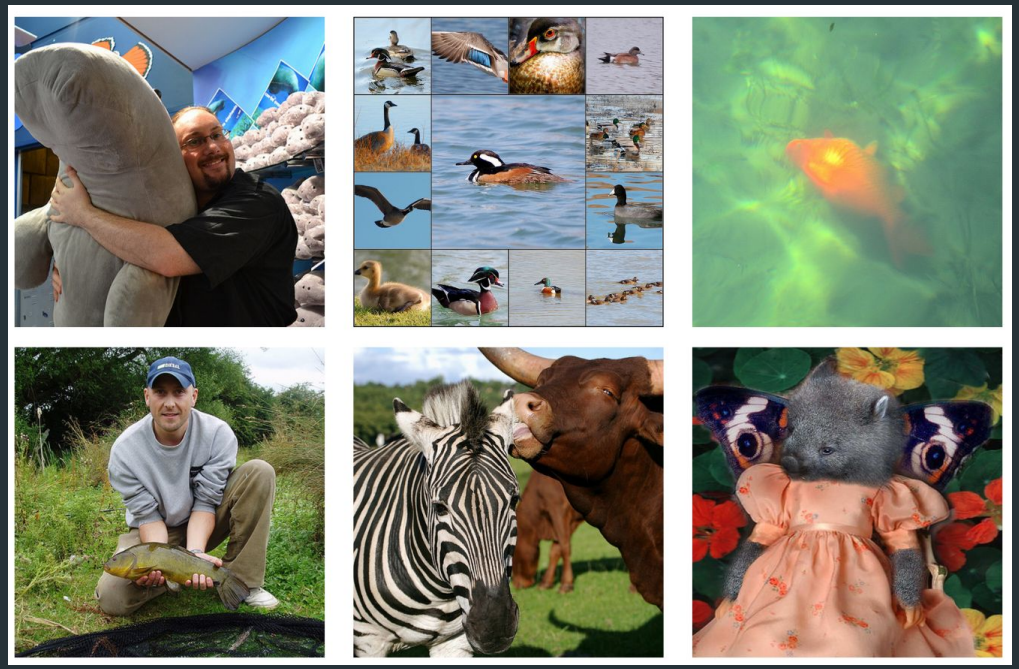
- ImageNet 1K: 269 classes of wild animals out of 1,000 total

BUGS IN THE DATA:
HOW IMAGENET MISREPRESENTS BIODIVERSITY

Alexandra Sasha Luccioni*
Hugging Face
sasha.luccioni@huggingface.co

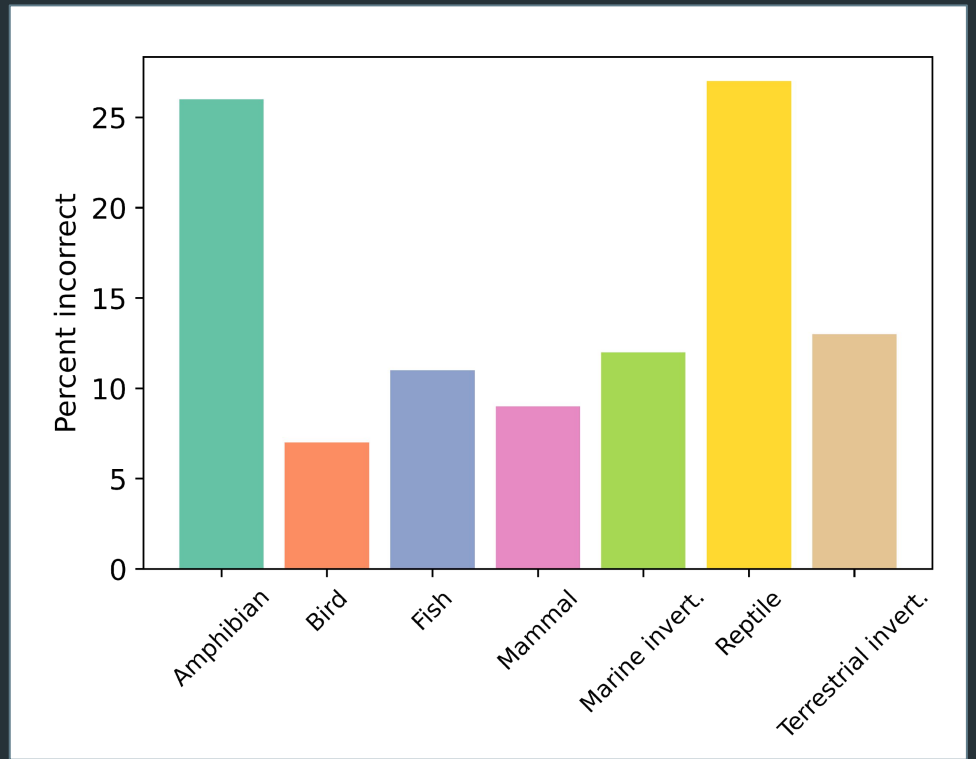
David Rolnick*
McGill University, Mila
drolnick@cs.mcgill.ca

Taxonomic group	# classes	# images
Amphibian	8	400
Bird	57	2850
Fish	16	800
Mammal	93	4650
Marine Invertebrate	20	1000
Reptile	35	1750
Terrestrial Invertebrate	40	2000
Total	269	13450



Representativeness- *Bugs in the Data (2023)*

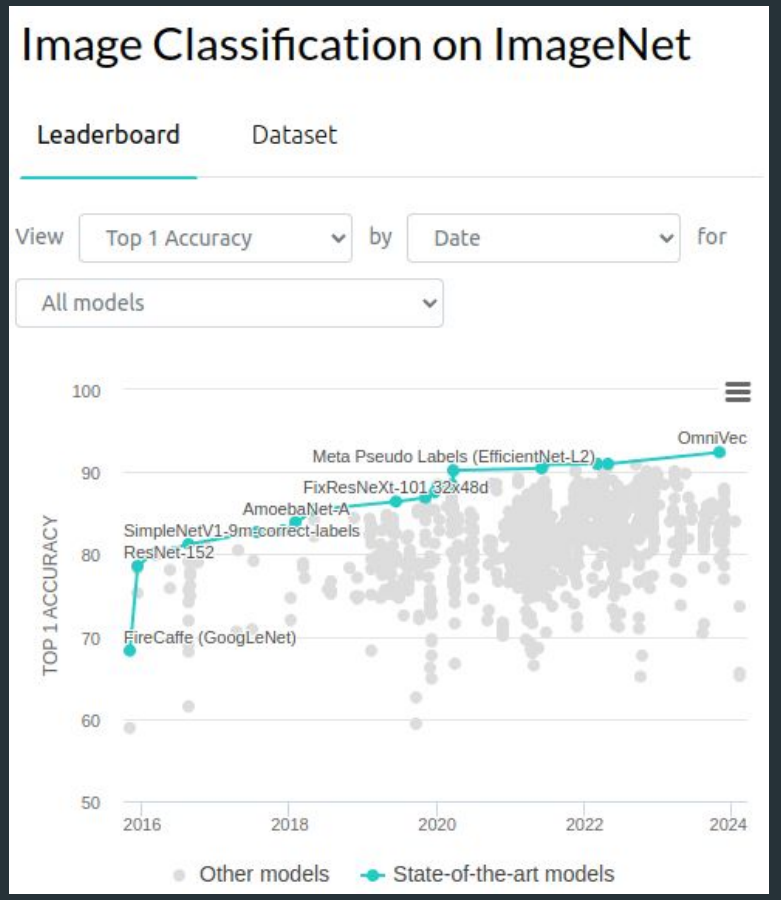
- Worked with domain experts in each of the taxonomic groups to verify images and classes
- 12% of images are incorrect, with some classes >90% wrong
- Sources of error:
  - Annotator inexperience
  - Prevalence effects (e.g. ferrets)
  - Nomenclature (e.g. ”leaf beetle”)
- This illustrates the **danger** of using data to train ML models for real-world tasks, assuming representativity



Class Label	Incorrect
black-footed ferret, *Mustela nigripes*	98%
rock crab, *Cancer irroratus*	96%
tailed frog, bell toad, *Ascaphus trui*	96%
kit fox, *Vulpes macrotis*	92%
goldfinch, *Carduelis carduelis*	90%
green lizard, *Lacerta viridis*	86%
night snake, *Hypsiglena torquata*	82%
green snake, grass snake	70%
mud turtle	62%
horned viper, cerastes, *Cerastes cornutus*	60%

Evaluation

- “You can’t improve what you can’t measure”
- And yet, AI leaderboards like Papers With Code tend to only measure performance-based metrics, e.g. accuracy
- Real-world constraints on model deployment that often result in trade-offs being made between different factors that include:
  - efficiency
  - robustness
  - fairness
  - data quality
- Evaluating **multiple aspects of system performance** and, importantly, clearly reporting the results of these evaluations, is key



Evaluation - AI Energy Star Ratings

The aim of this project is to develop an **Energy Star rating system** for AI model deployment that will guide members of the community in choosing models for different tasks based on energy efficiency.

⭐ **AI Energy Star Rating Project** ⭐

👫 **People involved** 👫:

@sasha, @yjernite, @IlyasMoutawwakil, @regisss, @sarahooker, Carole-Jean Wu,....

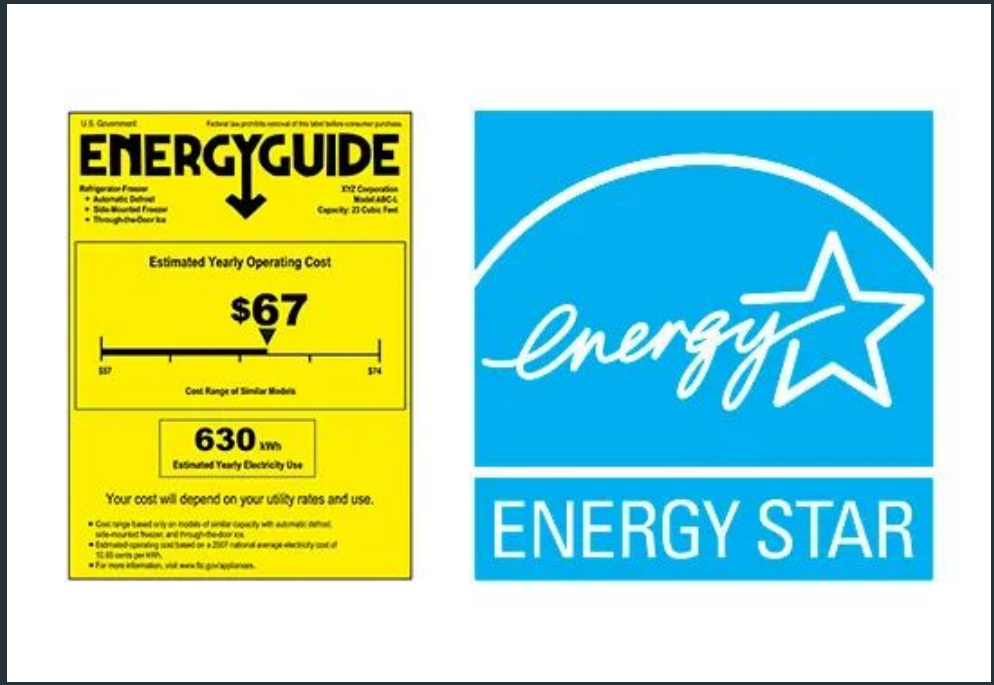
🌐 **Website:** Coming Soon!

🌎 **Project Mission** 🌎 **:**

To develop an Energy Star rating system for AI model deployment that will guide members of the community in choosing models for different tasks based on energy efficiency.

⏳ **Timeline (2024)** ⏳ **:**

mid-end Feb: project scoping and definition **early-mid March**: defining testing datasets and experimental protocol **mid-late March**: experimentation on HF cluster **April**: report writing + launch!



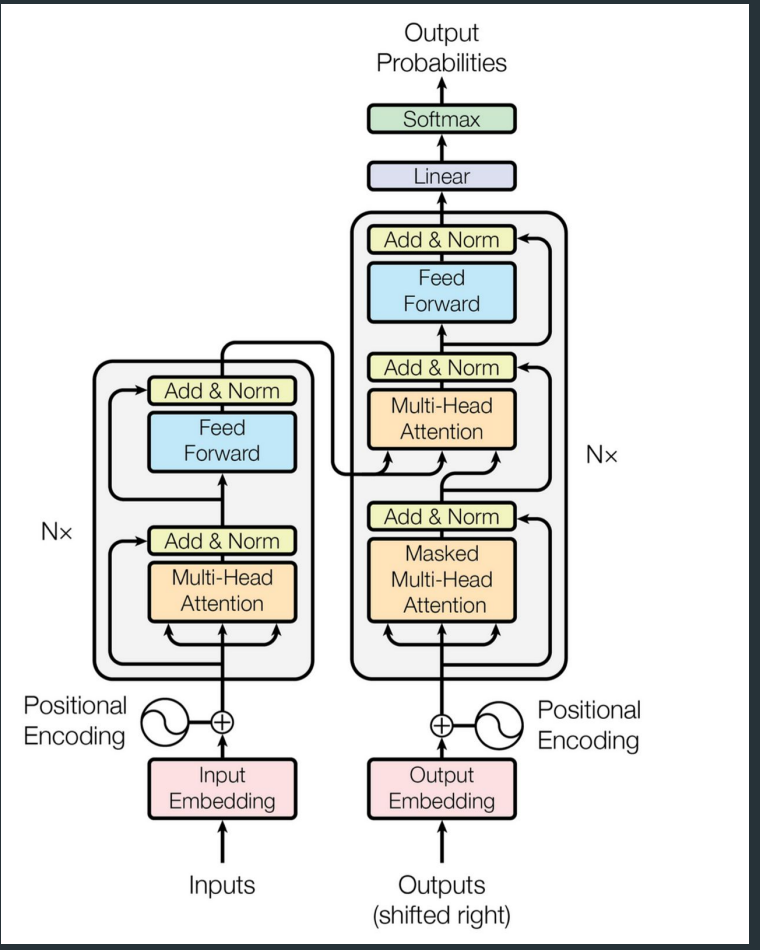
Please reach out if you want to collaborate on this!

Evaluating Rebound Effects

- Considerations around the the wider **rebound effect**s of proposed AI solutions is important.
- In the case of AI systems that improve aircraft efficiency, more efficient aircraft can result in cheaper airfare and therefore, more travel overall
- AI systems that carry out **precision agriculture** can replace already **underprivileged communities** such as migrant workers that traditionally harvested crops by hand.
- When a single metric is considered, these rebound effects don’t get measured and reported

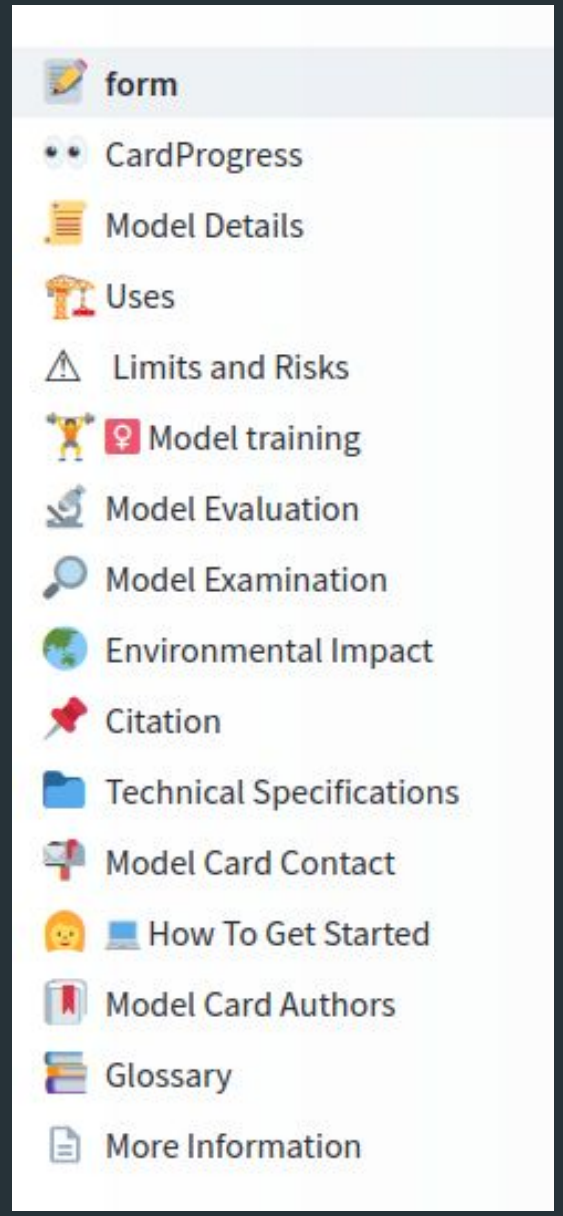
Transparency

- Modern AI systems are **not inherently transparent and interpretable** given the **complexity** of architectures they espouse and the **number of parameters** they contain
- While transparency is widely recognized as a **fundamental principle** in science in general, actualizing it in practice in AI can be challenging.
- Transparent research practices enable **replicability of results**, which in turn facilitates i**ndependent scrutiny**, **validation**, **further development of research** findings by other scientists
  - This comes with an added benefit for sustainability, since it **avoids the necessity** of having to **re-run experiments**



Model Cards and Datasheets

- In AI ethics, it is now a common practice **datasheets** and **model cards**, both of which are tools used to **understand** models and datasets
- However, while these artifacts have proven their utility, not including **discussions of carbon footprint** continues to silo these discussions and concerns.
- There is still an utter l**ack of information** about the **wider environmental impacts of AI** as a field

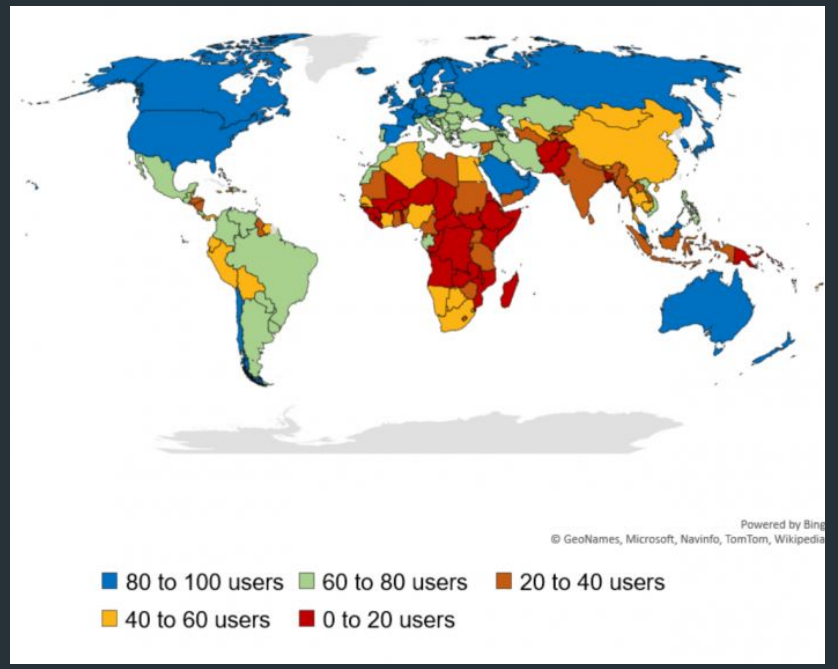


Social Transparency

- The notion of transparency in AI can be expanded to encompass "**social transparency**", which involves integrating socio-technical aspects in the description and understanding of AI systems
- Transparency can become a principle that enhances **ethical compliance** and promotes **environmental responsibility,** thus contributing to the broader goal of sustainability.

Equity

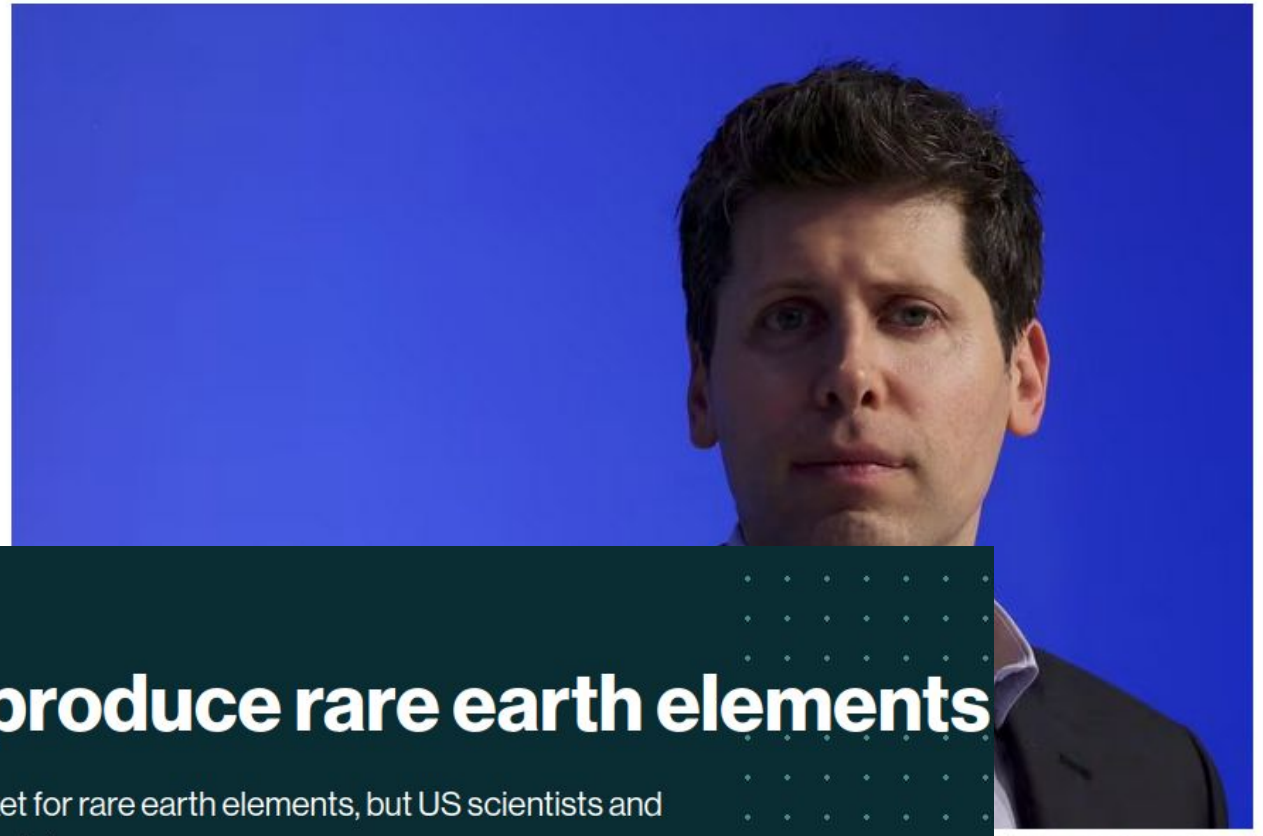
- Modern AI research and practice are **not equitable:**
  - its **cost** in terms of computer hardware as well as human skills means that only a small percentage of organizations can contribute
- With the recent advent of AI models of ever-increasing scale and complexity**, the digital divide in AI** is only increasing
- The majority of AI systems overlook issues of **justice** and **power**, focusing predominantly on the **climate-positive** aspects of technologies and not **who stands to benefit** from them.



OpenAI CEO Altman says at Davos future AI depends on energy breakthrough

Reuters

January 16, 2024 12:39 PM EST · Updated 17 days ago



co, California, U.S. November

BUSINESS

The race to produce rare earth elements

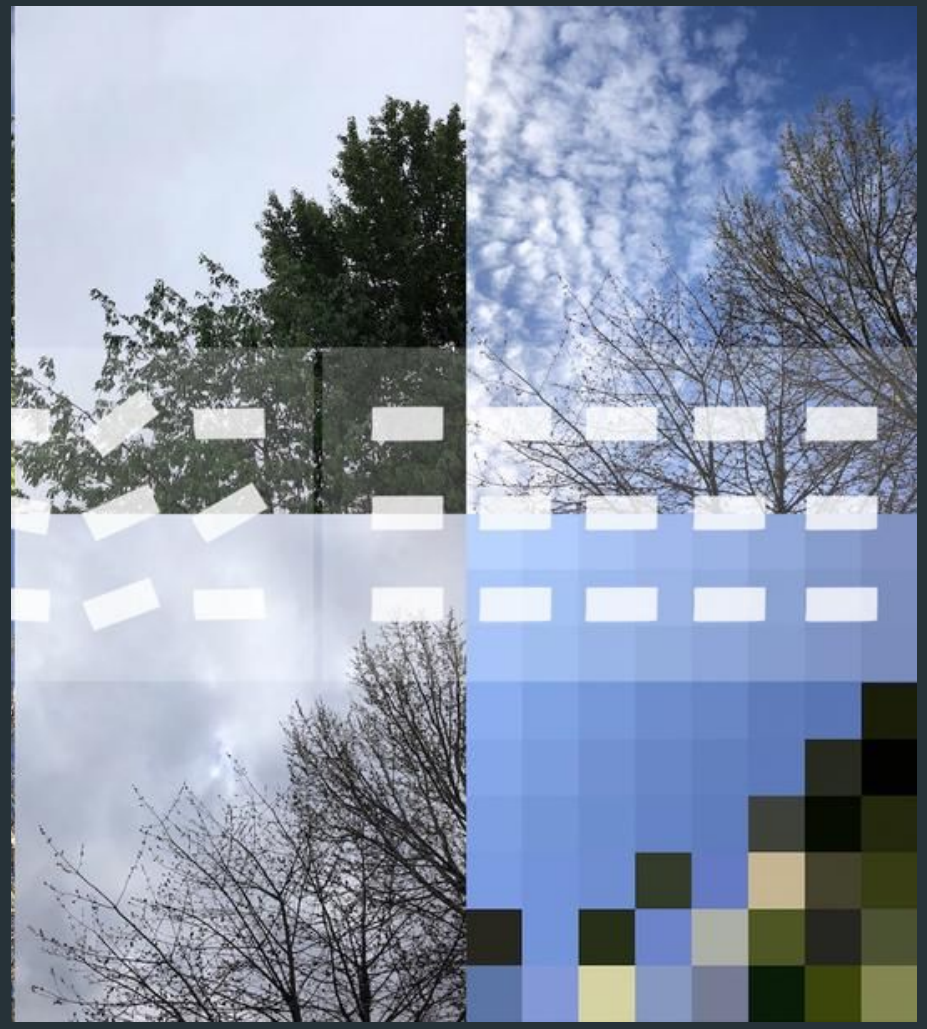
China has dominated the market for rare earth elements, but US scientists and companies are scrambling to catch up.

By Mureji Fatunde

January 5, 2024

Conclusion

- In the context of AI, there are multiple facets of AI technologies that have to be taken into consideration
- It is becoming increasingly clear that AI **ethics** and **sustainability** are interdependent: they must go hand in hand to ensure a holistic societal impact.
- Therefore, **integrating AI ethics with sustainability** is not just **beneficial** but **necessary** to increase their contribution towards sustainable advancement.



Thank you for your attention!

Image Credit:
Better Images of AI
(https://betterimagesofai.org/)