Issue #1 | Summer 2022

sustain

Sustainable AI in Practice

Artificial Intelligence: How To Make It More Sustainable

Overview:
Sustainability Criteria

Assessing the sustainability of AI based on 13 criteria

Green Data Centers:
Sustainable Radiators

Water-cooled servers prevent the energy needed for AI from going to waste

Standpoint:
Regulation

Political solutions are needed to address the risks posed by AI

sustain
Sustainability Index
for Artificial Intelligence

Project Partners:

ALGORITHM WATCH

iöw
INSTITUT FÜR ÖKOLOGISCHE WIRTSCHAFTSFORSCHUNG

DAI-Labor
TU Berlin

Supported by the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) based on a decision of the German Bundestag

Federal Ministry
for the Environment, Nature Conservation,
Nuclear Safety and Consumer Protection

Dear Reader,

One figure keeps popping up in the discussion on the sustainability of AI: The emissions generated during the research and development phase of large-scale language models are equivalent to the emissions of five cars throughout their lifecycle. This figure comes from a seminal analysis by Emma Strubell, Ananya Ganesh and Andrew McCallum. Of little help, however, are the many inaccurate or false statements that have since referenced this calculation – often making imprecise statements about the CO_2 costs of AI systems.

The discussion on the sustainability of AI deserves more accuracy, more nuance, more scrutiny and more evidence! The environmental, social and economic sustainability costs of AI urgently need to be addressed by academia, industry, civil society and policy makers – based on evidence. With the first edition of our SustAIn magazine, we hope to fuel this debate.

This magazine stems from our research project SustAIn in which we developed a framework to assess the sustainability of AI systems. Throughout this magazine, we demonstrate how more sustainable AI is already being achieved in practice. We need good, real word examples, methodological innovation and differentiated perspectives to decide what research and political actions we need to foster sustainable AI.

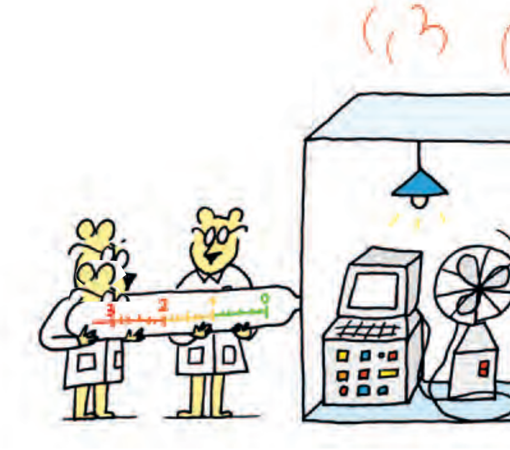
We hope you enjoy our magazine!

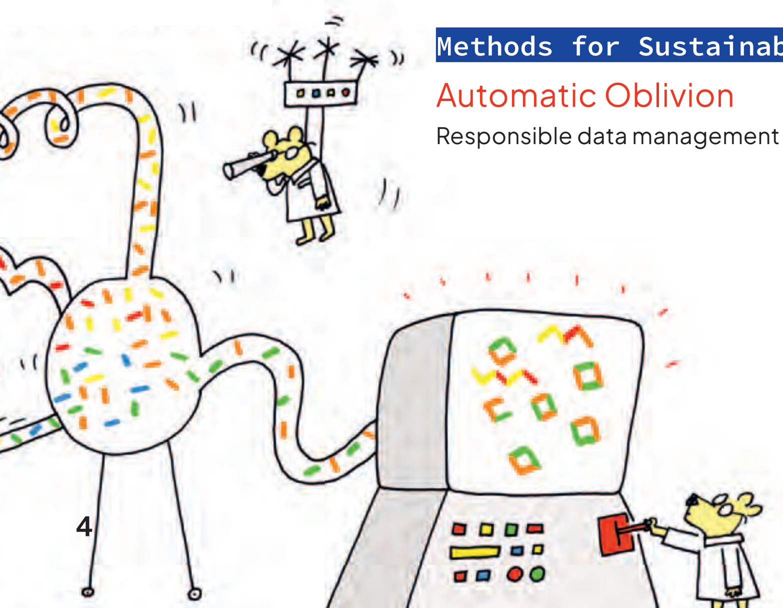
Dr. Anne Mollen

Project Manager “SustAIn:
The Sustainability Index for Artificial Intelligence”

AlgorithmWatch

CONTENTS





We need to build green date bases.
Mmm mm... right.
You won't regret it label
We want Real green labels!
NOW!
Sleepless night certified

sustain
Sustainability Index
for Artificial Intelligence

Project Partners:

AlgorithmWatch

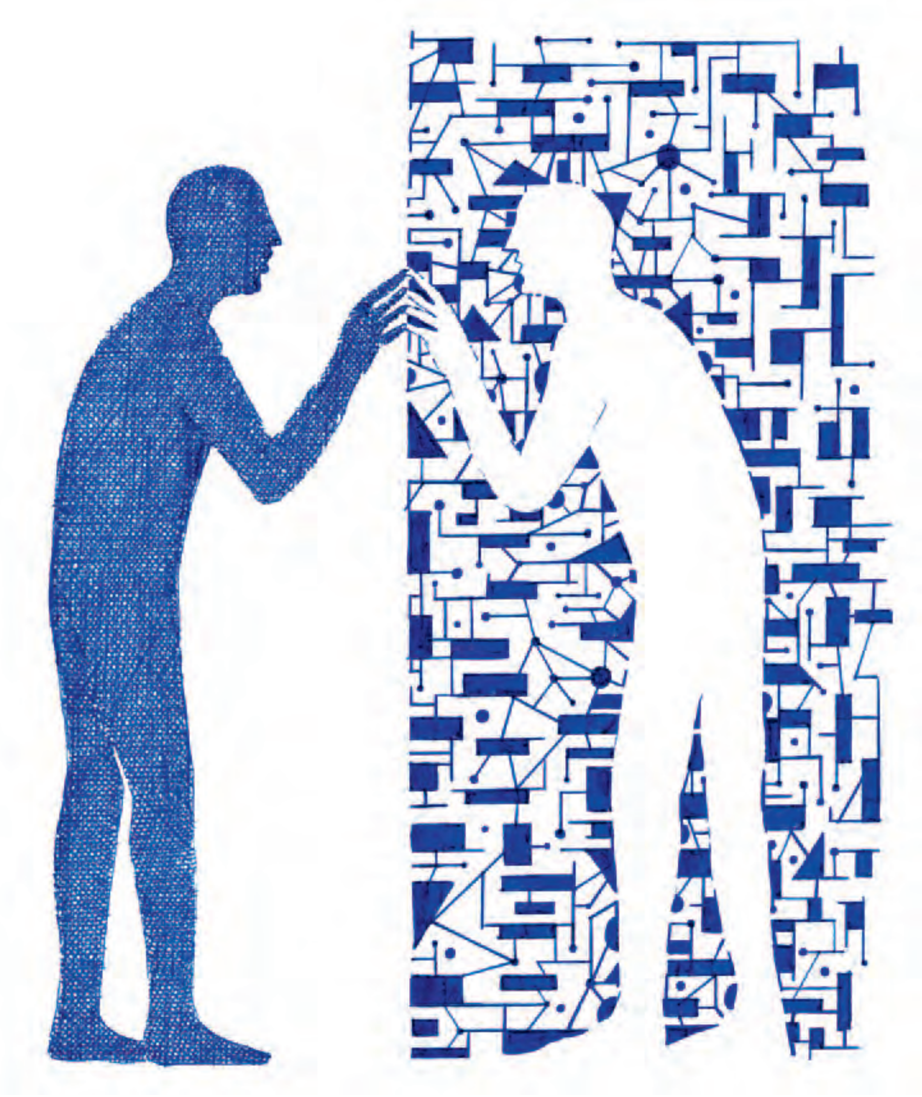
AlgorithmWatch is a non-profit research and advocacy organization whose aim is to monitor and analyze automated decision-making (ADM) systems and their impact on society. One pillar of our work is addressing the sustainability of ADM systems.

Institute for Ecological Economy Research (IÖW)

IÖW is a leading scientific institute in the field of practice-oriented sustainability research. We devise strategies and approaches for viable, long-term economic activity – for an economy which enables a good life and preserves natural resources. We have been dealing with issues of the future for more than 30 years, consistently finding new and frequently unusual answers.

Distributed Artificial Intelligence Laboratory

The DAI laboratory at the Technical University of Berlin views itself as a mediator between university-driven research and industrial applications. With our interdisciplinary team, we create innovations and translate university research into applications for everyday life in close cooperation with other scientific and industrial institutions.



SustAIn:

The Sustainability Index for Artificial Intelligence

SustAIn is an interdisciplinary team from AlgorithmWatch, the Institute for Ecological Economy Research and the DAI laboratory at the Technical University of Berlin that is looking for ways to define and measure the sustainability of AI. The project is funded by the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) as part of its initiative promoting AI lighthouse projects for the environment, climate, nature and resources.

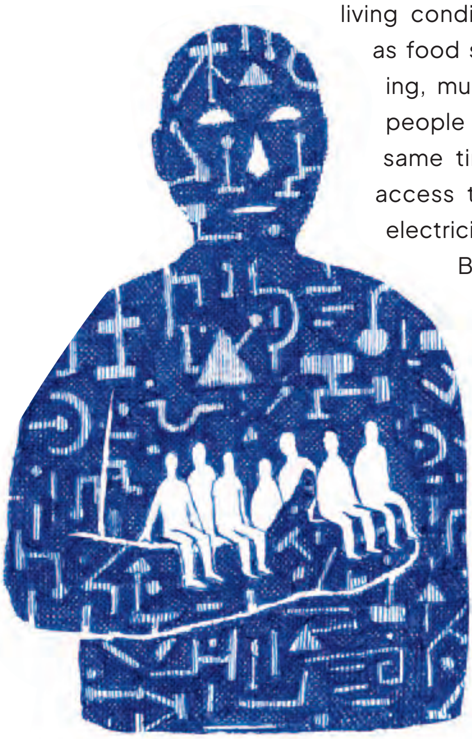
Practical applications of our sustainability criteria? Look for the info boxes throughout the magazine.

Assessing the Sustainability of AI

Sustainable AI respects planetary boundaries, it doesn't exacerbate problematic economic dynamics and it doesn't threaten social cohesion. As part of the SustAIn project, we have used this as the basis for defining 13 criteria that organizations should consider in order to develop and use AI more sustainably.

Social Sustainability

The socially sustainable development and application of Artificial Intelligence emphasizes people, society and just living conditions. Basic needs, such as food supply or adequate housing, must be met to ensure that people live a dignified life. At the same time, they must also have access to infrastructure, such as electricity, water and the internet. Beyond that, though, it is also a matter of ensuring societal cohesion. The human rights of particularly disadvantaged groups in society must be protected in the digital space as well. A socially sustainable society allows its people to develop freely. The empowerment approach of Nobel laureate economist Amartya Sen and moral philosopher Martha Nussbaum holds that the sustainable development of society must offer people opportunities for fulfillment. They must be able to draw on a wellspring of material and cultural resources to help them exercise their rights.



AI systems must also do their part to uphold human dignity. They must not exclude, disadvantage or discriminate against anyone and must not restrict human autonomy and our freedom to act. Values such as fairness, inclusion and freedom must be factored into the design, development and application of AI. In particular, the ability to think, reason and act in a human way also must not be limited by systems.

Transparency and Accountability

Those who use or interact with AI should be informed in advance that AI is being used and must be able to understand the outcomes that result. As such, critical information about AI systems must be disclosed and responsibility for their output must be clarified.

Non-Discrimination and Fairness

There should be an awareness of fairness in the development and application of AI. In addition, AI systems should be regularly reviewed for possible discrimination.

Technical Reliability and Human Supervision

Weak points in AI systems should by systematically identified through risk assessments. In addition, high data quality must be guaranteed and human intervention into the systems should be possible.

Self-Determination and Data Protection

Small data sets, encryption, the right to object to the use of personal data, and other measures strengthen informational self-determination and data privacy.

Inclusive and Participatory Design

End users, the people affected and other stakeholders should be involved in the AI design process. Teams planning and developing AI should be diverse and interdisciplinary.

Cultural Sensitivity

The context of the AI application must be considered during its development. AI systems should therefore be adaptable and re-trainable when used in different cultural contexts. In particular, local knowledge and data sets should be utilized.

Environmental Sustainability

AI is often seen as an important tool for addressing the climate crisis. The potential applications of AI systems are wide-ranging: They are designed to make resource consumption more efficient, bring about more efficient transportation and more effective urban planning, create a more sustainable energy system and even facilitate research into new materials. Rare earths, necessary for chips and circuit boards, must be mined at great expense. The finished hardware also has to be transported to the site where it will be used. Energy is consumed at data centers to develop and deploy AI systems, and at the same time, IT infrastructure must be cooled to protect components from damage. And the hardware also must be replaced regularly, which in turn leads to new material requirements and a large amount of electronic waste. In the worst-case scenario, improper disposal can also release harmful chemicals.

The aim of environmental sustainability is to preserve nature in order to keep our planet inhabitable for future generations. The "planetary boundaries" developed among others by Johan Rockström define thresholds that, if exceeded, would result in irreversible environmental damage. AI systems impact many of these boundaries, either directly or indirectly. AI systems are the opposite of environmentally sustainable if they consume more resources than are saved or even reproduced through their use. In addition to the material consumption for hardware, their immense energy consumption and the associated emissions are an obstacle on the road to environmental sustainability.

Energy Consumption

Energy efficiency should be monitored during AI development and, if necessary, optimized through appropriate methods such as model compression.

CO_2 and Greenhouse Gas Emissions

CO_2 efficiency can be increased through the use of a sustainable energy mix, the appropriate choice of time and location for training, and by offsetting the CO_2 emissions generated.

Sustainability Potential in Application

AI systems can have a sustainable impact if they take sustainability into account in their decision-making – if, for example, they promote sustainable products or they minimize the broader consumption of resources.

Indirect Resource Consumption

The hardware necessary for AI systems requires additional energy and resources. Here, it is critical to consider environmentally friendly production and disposal.

Economic Sustainability

Economic sustainability expands the horizon of economic activities: Rather than focusing the economy only on satisfying the needs of people living today, an economically sustainable perspective also plans for meeting the needs of humanity in the future. This change in consciousness is urgent against the backdrop of the "Grand Challenges" of climate change, ongoing loss of biodiversity and species, and the growing scarcity of resources. In the course of a socio-ecological transformation, fundamental questions of fairness arise because production and consumption cycles determine whether the distribution of natural resources can be reconciled with a decent and self-determined life. Economic sustainability embeds the economy between social and environmental guard rails. The effects of AI systems must also be viewed in this context. Systems that have far-reaching effects on the distribution of wealth in society (for example, in the allocation of social benefits, loans or housing) as well as on economic structures and dynamics must be used in a particularly responsible manner.

Market Diversity and Exploitation of Innovation Potential

To prevent conentration in AI markets, fair access must be established for AI development through, for example, open data pools, open source code or even interfaces (APIs).

Distribution Effect in Target Markets

Access to AI applications is not available to all economic actors, leading in some cases to competitive distortions or even market concentrations. Inclusivity could be expanded by enabling models to work with small sets of data, or by enabling small and medium-sized companies to use AI through funding opportunities.

Working Conditions and Jobs

Fair working conditions should be ensured along the entire value chain of AI development. If AI is deployed in the work-place, the impact on workers should be assessed in advance and, where necessary, compensated for.

We have differentiated and operationalized these criteria into more than 40 indicators. They can be found in the discussion paper:

Rohde, F., Wagner, J., Reinhard, P., Petschow, U., Meyer, A., Voβ, M., & Mollen, A. Sustainability criteria for artificial intelligence. A Text series of the IÖW, 220, 21.

Download: https://www.ioew.de/fileadmin/user_upload/BILDER_und_Downloaddateien/Publikationen/2021/IOEW_SR_220_Nachhaltigkeitskriterien_fuer_Kuenstliche_Intelligenz.pdf (in German)

ANDREAS MEYER

… is a research associate at the Distributed Artificial Intelligence Laboratory at TU Berlin, where he is researching applications of Machine Learning methods for load forecasting and the sustainability of AI systems.

ANNE MOLLEN

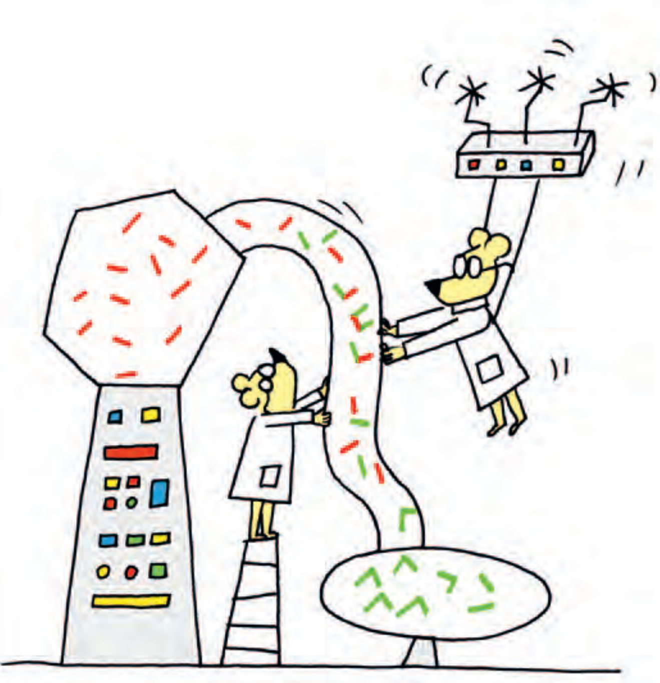
… is senior policy and advocacy manager at AlgorithmWatch. She works on the sustainability of automated decision-making (ADM) systems. Other areas of focus include ADM systems in the workplace and in the public sector.

FRIEDERIKE ROHDE

… conducts research into sustainability at the Institute for Ecological Economy Research (IÖW). She is completing her Ph.D. at the TU Berlin and works on socio-technical futures in the context of digital transformation, social in-novations and algorithmic decision-making systems.

JOSEPHIN WAGNER

… is a research associate at the Institute for Ecologi-cal Economy Research. In the research field of environmental economics and policy, she focuses on the topics of digitaliza-tion and social change as well as economic and institutional analysis of environmental policies.

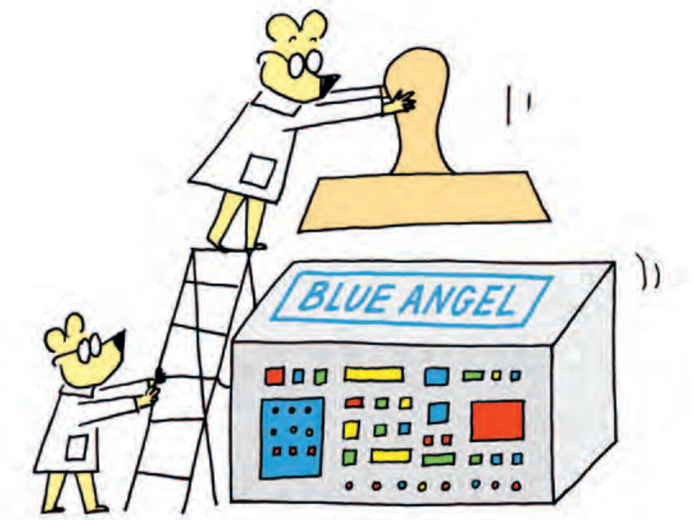


Sustainable AI 101

Certain terms keep coming up in discussions on sustainable AI: *foundation models* are particularly energy-intensive, *model cards* can create more transparency, *APIs* promise openness ... We explain some of the central terms here.

APIs

AI systems or platforms can provide interfaces (APIs = application programming interfaces) for third-party vendors and other AI-developing companies to augment AI systems with external services. APIs facilitate the transfer of data and communication between two programs. Developers can use them to search for, collect or share data, or to build and adapt functions into their own software products. A basic distinction is made between private APIs, partner APIs and public APIs. Private APIs can only be accessed by internal developers and workers within a single company. Restricted APIs, called partner APIs, can only be used by select companies under certain contractual requirements. Public APIs are freely available and can be used by any company without restrictions. Unlike open-source software, however, public APIs do not provide insight into the source code or allow for free customization.
Public APIs are particularly beneficial to small and medium-sized enterprises and NGOs, which often lack the resources and expertise to develop their own competitive AI software. Public APIs allow them to make processes more efficient and resource friendly and to flexibly integrate AI functions into their own software.



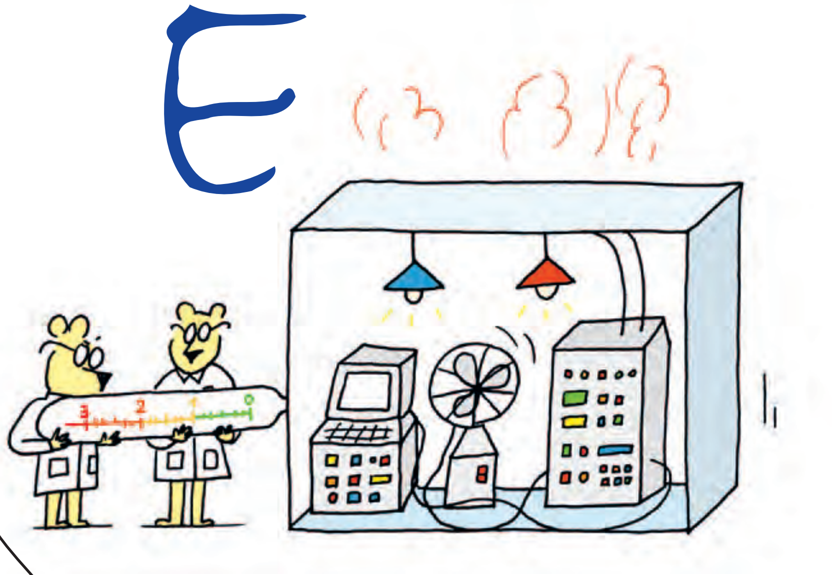
Certification of Hardware and Data Centers

The amount of energy required to operate data centers is constantly growing. In German financial center Frankfurt, data centers are today responsible for 20 percent of all electricity consumption, and the trend is growing. Certifications can be helpful in terms of finding more sustainable alternatives when selecting data centers and hardware. The certificates are awarded for individual hardware components (for example, the Blue Angel or Energy Star labels for servers and data storage products), but also for entire data centers and colocation data centers, where companies rent space to operate their own IT hardware. In addition to the Blue Angel, there is also, for example, the CEEDA certificate for data centers.
The Blue Angel label certifies efficient power supply and air conditioning as well as the use of energy-efficient, durable and environmentally friendly hardware components. Consistent monitoring of data center operations and an annual report ensure ongoing compliance.

Data Pools

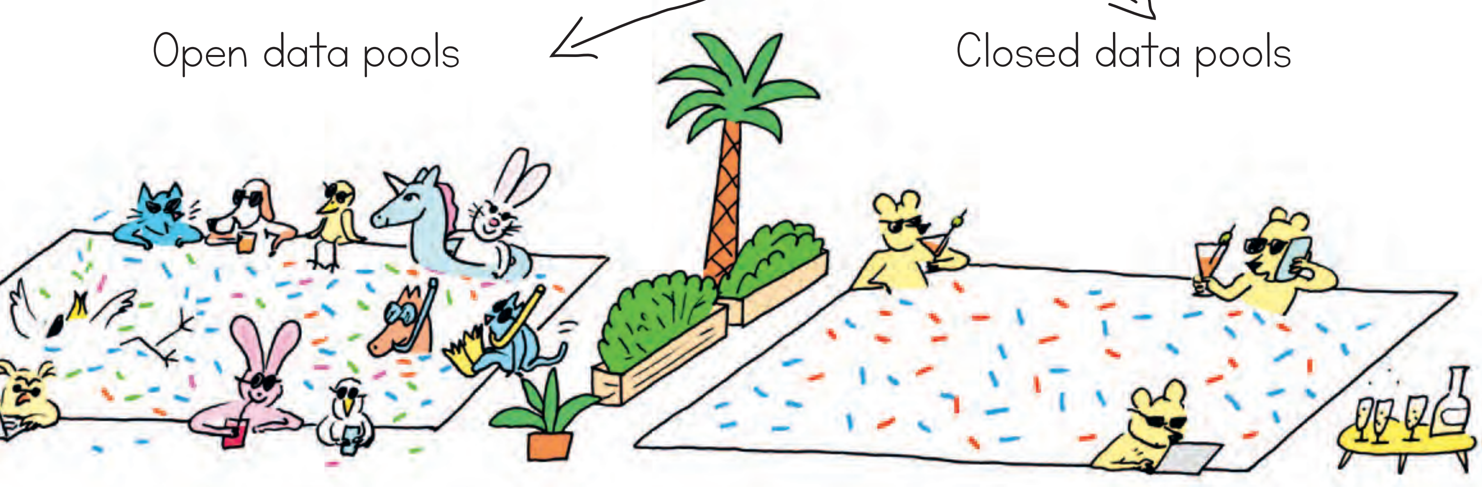
Data pools are datasets that are either self-generated by AI or come from internal or external sources. A distinction is made between open and closed data pools. In the case of open data pools, several companies share access to the data after agreeing on specific framework conditions regarding the use and customization of the pool. Closed data pools are data sets to which only one company has access. These data pools are advantageous in that companies don't have to share competitive advantages associated with certain data with their competitors.

From the perspective of economic sustainability, closed data pools are a delicate issue. They lead to so-called lock-in effects – which means that users are bound to a specific product or its provider in ways that make it difficult for them to switch to other products or providers due to unavoidable barriers such as high switching costs. They can also result in increased market concentration and even monopolies. The innovative strength of the market suffers as a result, and market diversity is severely constrained as single, large companies disproportionately benefit from AI. At the same time, though, closed data pools and the competitive advantages they provide incentivize companies to engage in practices that distort competition when obtaining data.



Efficiency Metrics for Data Centers

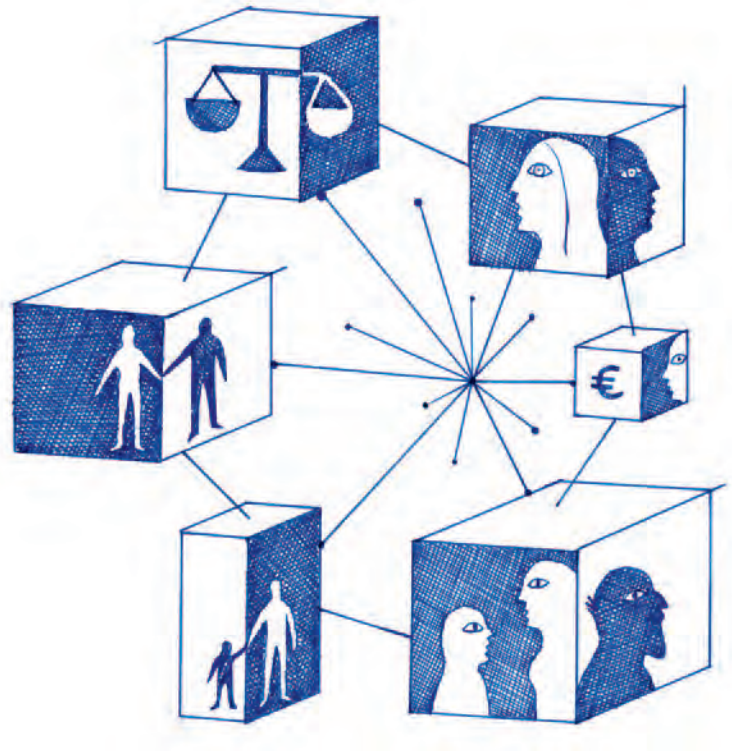
Efficiency metrics have been developed to measure the efficiency of data center operations and to identify existing weaknesses in different areas. The respective metrics map different types of resource consumption. Power Usage Effectiveness (PUE) is one of the most widely used efficiency metrics. It measures the efficiency of energy use and the relationship between total energy consumption and that of IT infrastructure. With a PUE of 1, all the energy expended would flow into the infrastructure. A value of 2 would mean that cooling, lighting and the facility itself require just as much power as the IT infrastructure. Anything that doesn't directly serve the operation of computing is considered non-infrastructure. In 2020, data centers in Germany had an average PUE of 1.63. The efficiency with which other resources are used can be measured in the same way. *Water Usage Effectiveness* (WUE) or *Carbon Usage Effectiveness* (CUE) record the amount of water consumed or carbon dioxide emitted for a fixed level of energy consumption. Many such efficiency metrics exist, also measuring things like the utilization of the technical infrastructure, for example, or the reuse of the waste heat generated.

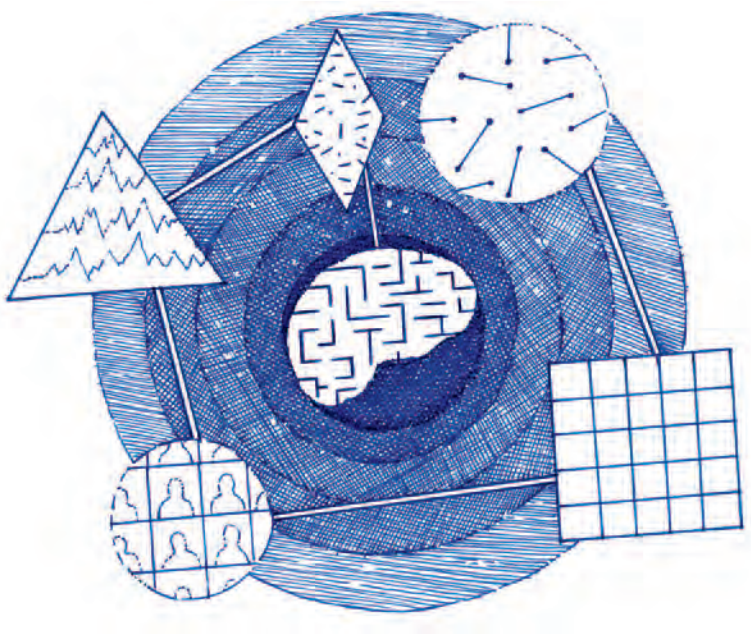


F

Fairness Metrics

Algorithms can discriminate against people for instance based on age, gender or skin color if, for example, the data used to train the models contains a bias and thus reproduces social prejudices. The fairness of models is measured and tested to avoid such discriminatory effects. The various statistical approaches used to do so are called fairness metrics. These metrics can be used, for example, to measure the likelihood of favorable decisions by the algorithm for groups with different demographic characteristics, such as age or income, or to test whether the accuracy of the model is the same for different subgroups (whether credit scores show a significant variation between males and females, for example). The online library *Tensorflow* provides an extensive list of the different methods available for measuring fairness (https://github.com/tensorflow/fairness-indicators). There are, however, limits to the measurability of fairness given that many of the influencing variables are difficult to quantify. That is why it is important to clearly define fairness and designate groups worthy of protection based on protected attributes such as ethnicity, origin and skin color.





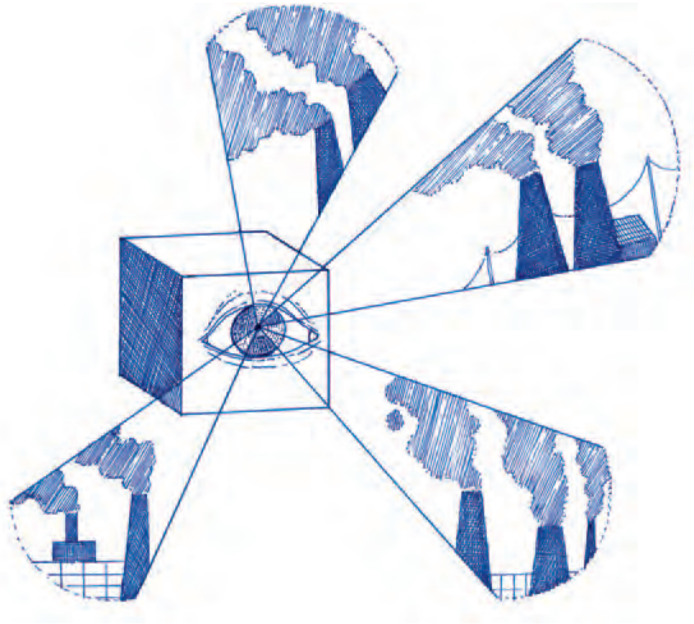
F

Foundation Models

A research group at Stanford University first coined the term foundation models in response to an advancing paradigm shift in the field of artificial intelligence. AI models for speech processing have been growing larger and more powerful for some time. Examples from recent years include OpenAI's GPT-3 and Google's BERT. In an often extremely complicated process, AI models learn pre-defined categorizations through the processing of previously annotated data. Called supervised learning, this process was used in areas like the recognition of objects within images. In what is known as self-supervised learning, basic models such as GPT-3 and BERT first attempt to identify and learn general patterns in the data. Since data doesn't have to be manually annotated for this purpose, the models can draw on significantly larger data sets and recognize complex relationships. These models can be used much more flexibly and diversely as a basis for various applications. These applications, however, also take on the characteristics and possible distortions inherent in the models. As the complexity of AI systems increases, so do the costs of developing them. As a result, the number of players capable of developing such large AI models is decreasing, leading to the increasing concentration of AI development.

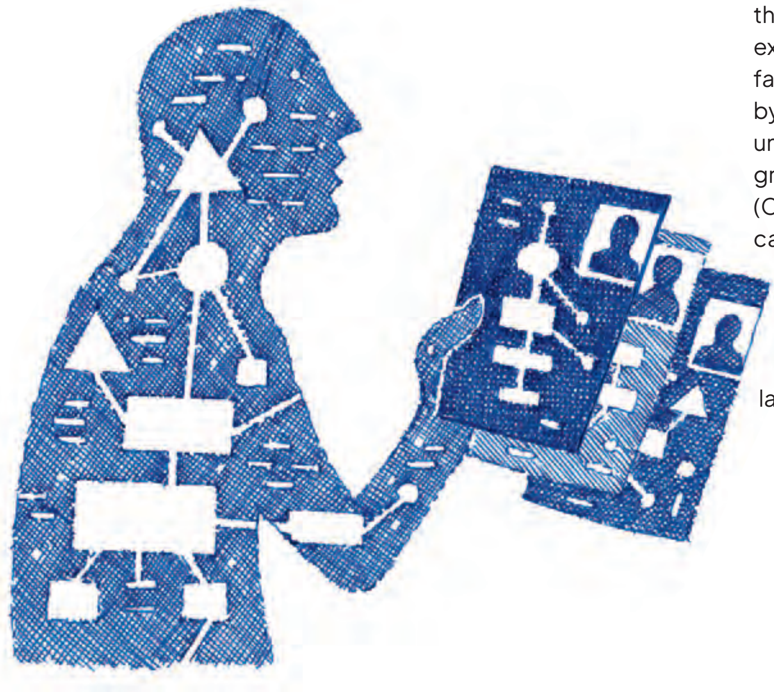
Model Cards

Model cards are used to document Machine Learning models. They are intended to provide information about the context in which algorithmic decision systems are used. The short documents detail the performance characteristics of the respective algorithm in a structured manner and also contain information about the context in which model training took place, including information about different cultural, demographic or phenotypic groups (ethnicity, geographic locations, gender or Fitzpatrick skin type, for example) and intersectional groups (such as age *and* ethnicity or gender *and* Fitzpatrick skin type). Complete documentation should also include the type and details of the Machine Learning model as well as the intended use and possible influencing factors. In addition, test and training data should be recorded on model cards along with any ethical issues or concerns. The aim of the documentation is to ensure that the Machine Learning behind the models in question becomes more transparent and comprehensible. Google, for example, has published a model card for an algorithm that recognizes faces in photos and videos. Another example is the model card for the commonly used language processing model BERT, published on the developer platform *Hugging Face* (https://huggingface.co/distilbert-base-uncased).



Tools for Measuring CO_2 Consumption of AI Applications

Over their life cycles, AI systems cause both direct and indirect CO_2 emissions. On the one hand, greenhouse gases are generated by the hardware required during raw material extraction, production, transport and disposal. These emissions, however, cannot be precisely quantified in most cases due to a lack of information available and the complexity of their supply chains. It is also difficult to directly link them to specific AI models. Most of the CO_2 emissions caused by AI systems are the product of the electricity consumed by the hardware when the systems are developed and deployed. The extent of the emissions depends mainly on two factors: the amount of electricity consumed by the hardware and the CO_2 intensity of the underlying power mix. Some tools can be integrated into the source code of the applications (CodeCarbon, experiment-impact-tracker, carbontracker) to determine the emission value. It is also possible to make projections regarding the expected emissions by considering the system specifications (with the Machine Learning Emissions Calculator or with carbontracker, for example).



Sustainable AI: Facts & Figures

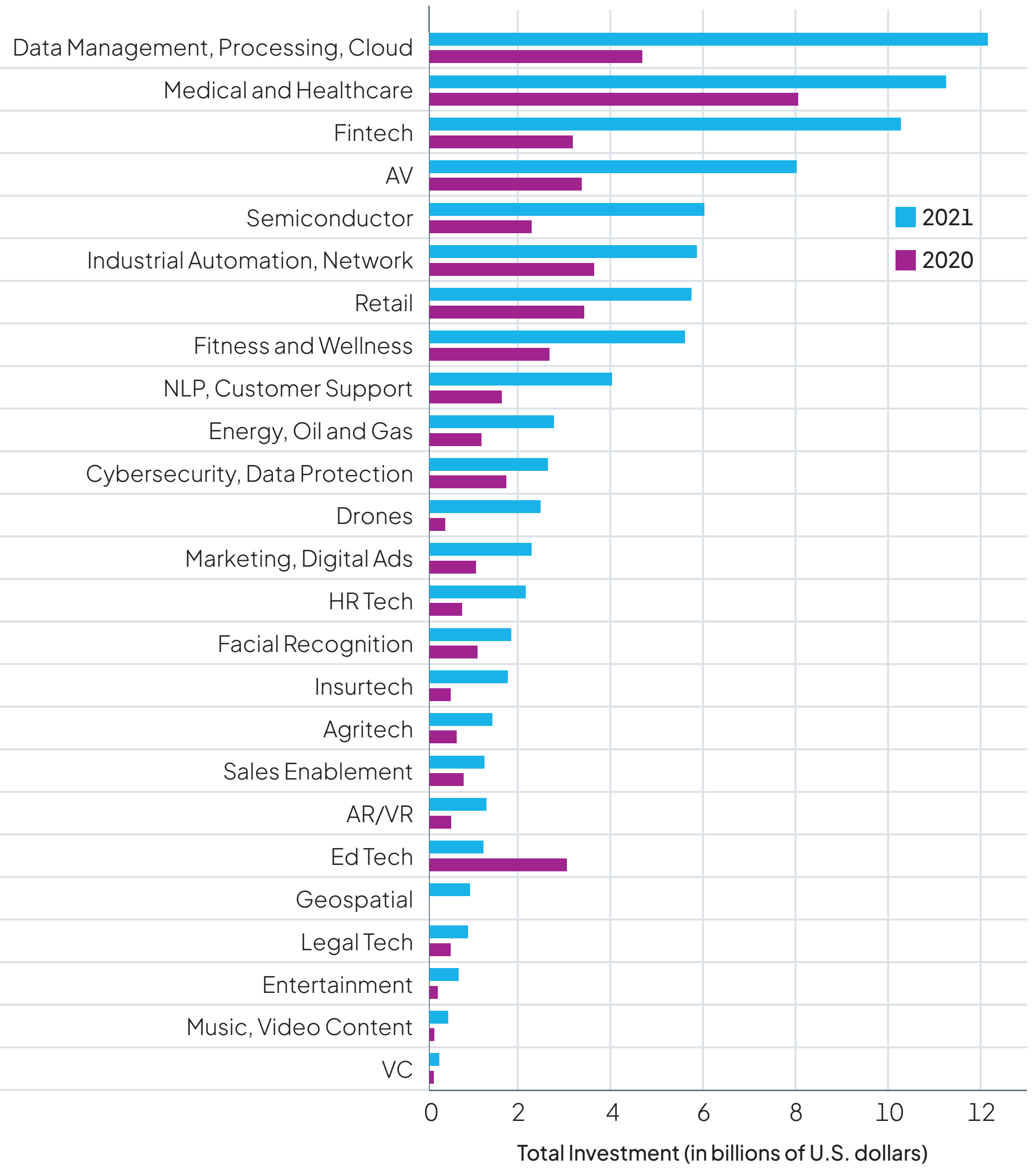
Stanford University's annual AI Index Report tracks, collates, distills and visualizes data related to Artificial Intelligence. Its aim is to provide a more thorough and nuanced understanding of the field of AI. We present some of the data from this year's edition that is relevant from a sustainability perspective.

Growth: Private AI Investment

The AI industry is growing rapidly, especially investments in data management, processing and clouds. In 2021, they increased by more than two and a half times compared to the previous year and amounted to around USD 4.69 billion. Two of the four largest private investments in 2021 went to data management companies.

PRIVATE INVESTMENT IN AI BY FOCUS AREA, 2020 VS. 2021

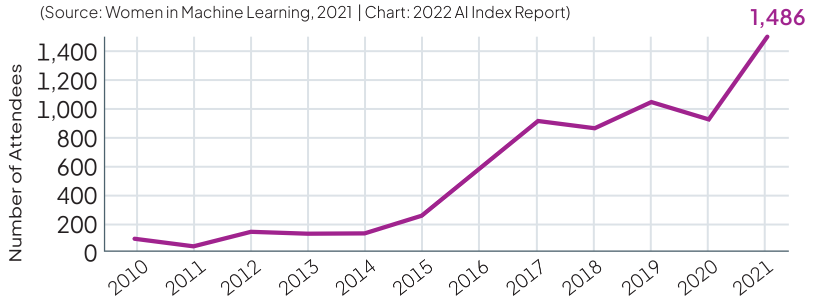
(Source: NetBase Quid, 2021 | Chart: 2022 AI Index Report)



Diversity: Women in Machine Learning

Diversity in AI is key against discrimination. Founded in 2006, Women in Machine Learning (WiML) is an organization dedicated to supporting and increasing the impact of women in Machine Learning. This data illustrates the number of attendees at WiML workshops over the years at NeurIPS – one of the most important AI and ML conferences.

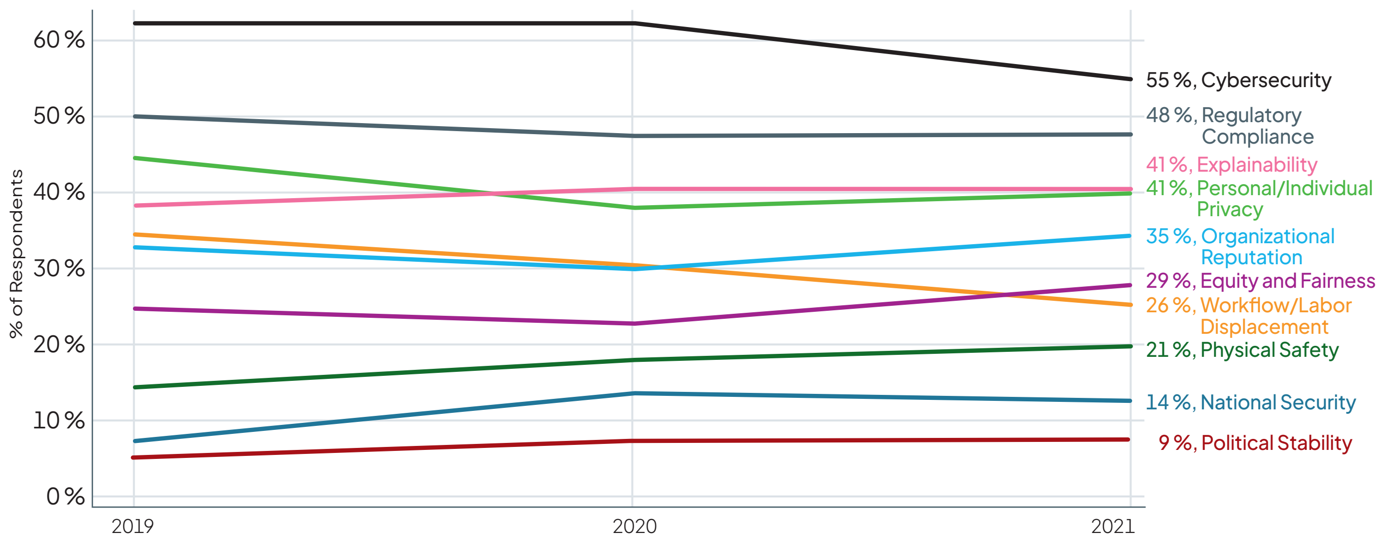
ATTENDANCE AT NEURIPS WOMEN IN MACHINE LEARNING WORKSHOP, 2010–2021
(Source: Women in Machine Learning, 2021 | Chart: 2022 AI Index Report)



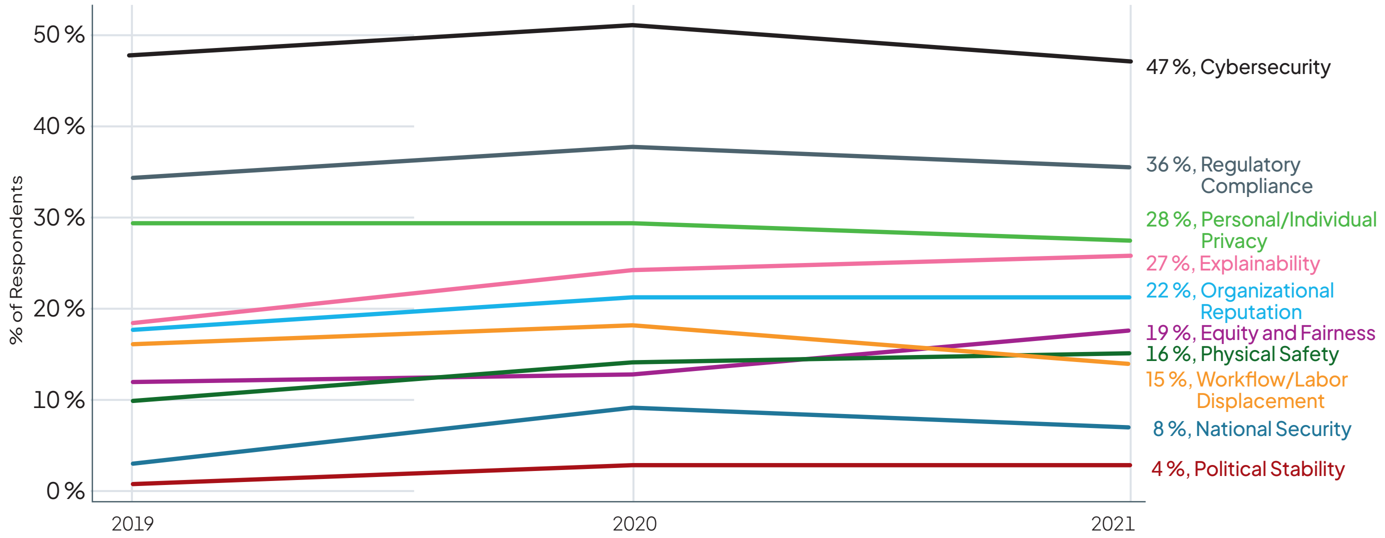
Risks: Consideration and Mitigation of Risks from Adopting AI

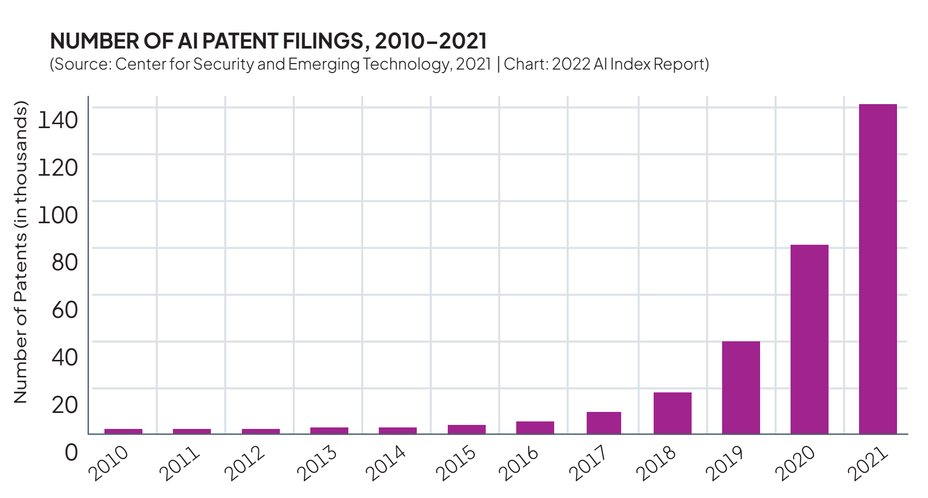
Risks not only need to be recognized, but also to be actively addressed. Currently, gaps remain between recognizing risks and acting upon them — a gap of 10 percentage points regarding risks relating to equity and fairness (29 percent to 19 percent), 12 percentage points for regulatory compliance (48 percent to 36 percent), 13 for personal/individual privacy (41 percent to 28 percent), and 14 for explainability (41 percent to 27%).

RISKS FROM ADOPTING AI THAT ORGANIZATIONS CONSIDER RELEVANT, 2019–2021
(Source: McKinsey & Company, 2021 | Chart: 2022 AI Index Report)



RISKS FROM ADOPTING AI THAT ORGANIZATIONS HAVE TAKEN STEPS TO MITIGATE, 2019–2021
(Source: McKinsey & Company, 2021 | Chart: 2022 AI Index Report)



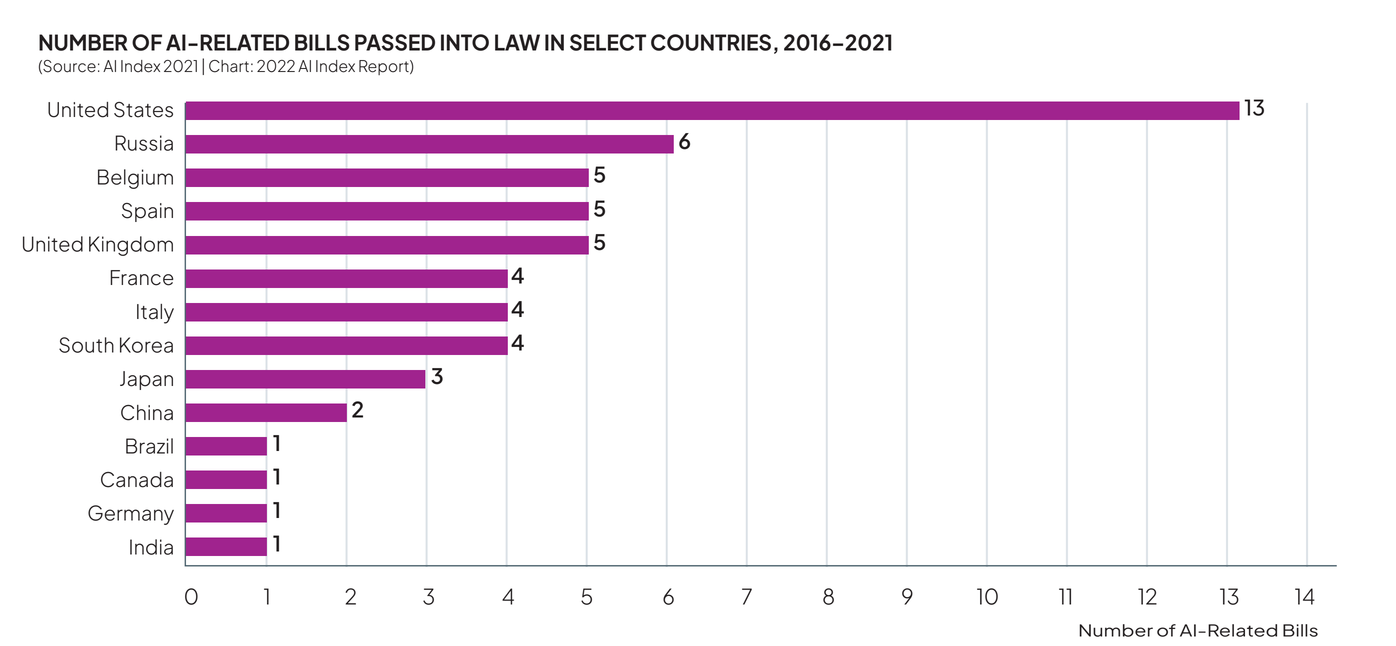


Growth:
AI Patents

AI is becoming more pervasive: The number of patents filed in 2021 is more than 30 times higher than in 2015, showing a compound annual growth rate of 76.9 percent.

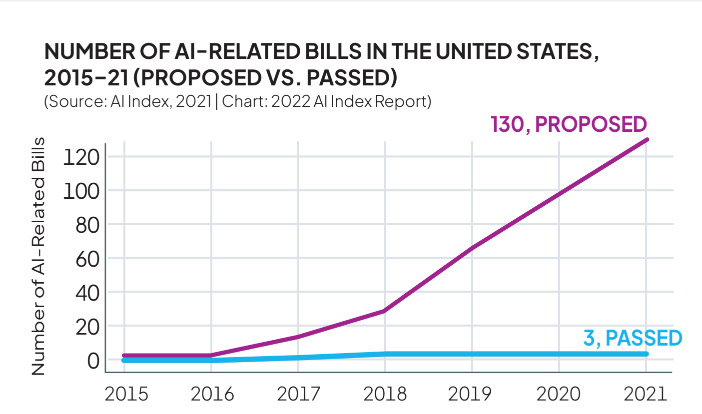
Governance:
Global Legislation Records on AI

Governments and legislative bodies across the globe are increasingly seeking to pass laws to regulate the development of AI. An analysis of laws passed in 25 countries by their legislative bodies that contain the words "Artificial Intelligence" showed that, taken together, a total of 55 AI-related bills have been passed.



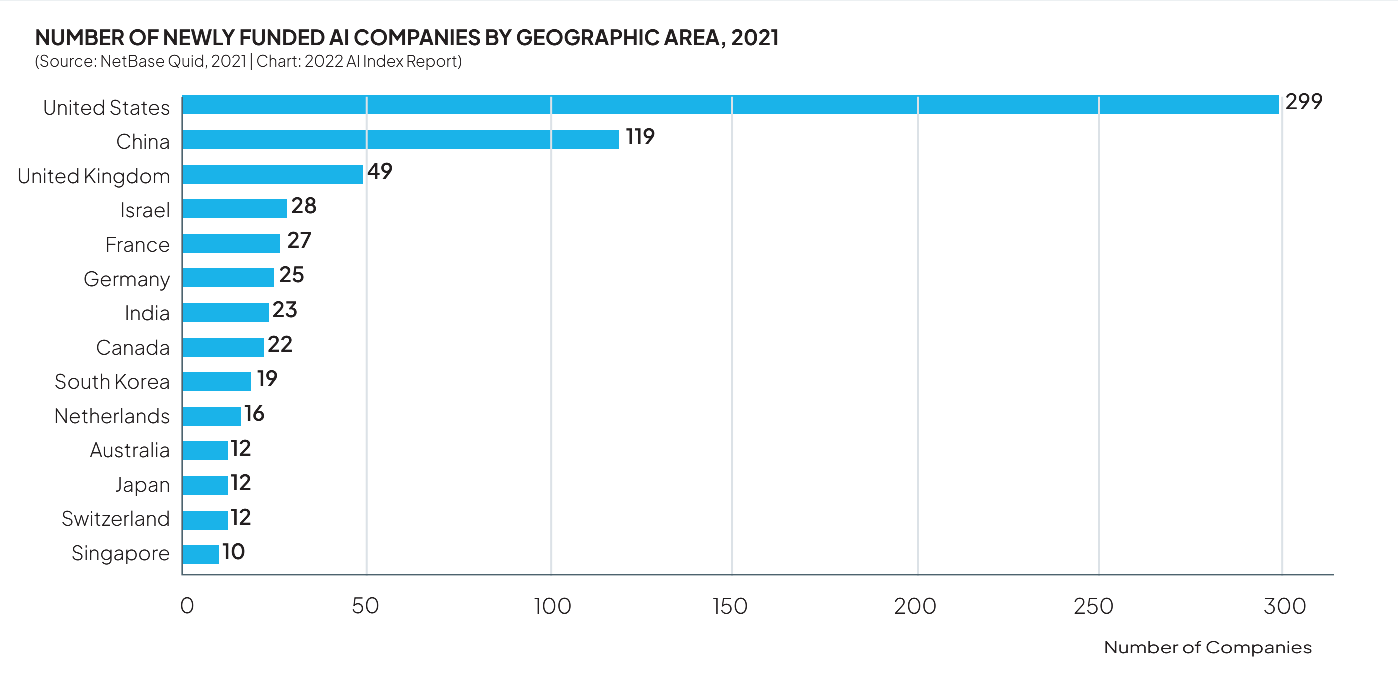
Governance:
Federal AI Legislation in the United States

The federal legislative record in the United States shows a sharp increase in the total number of proposed bills that relate to AI. In 2015, just one federal bill was proposed, while in 2021, there were 130. The number of bills related to AI being passed has not kept pace with the growing volume of proposed AI-related bills. This gap was most evident in 2021, when only 2 percent of all AI-related bills were ultimately passed into law.



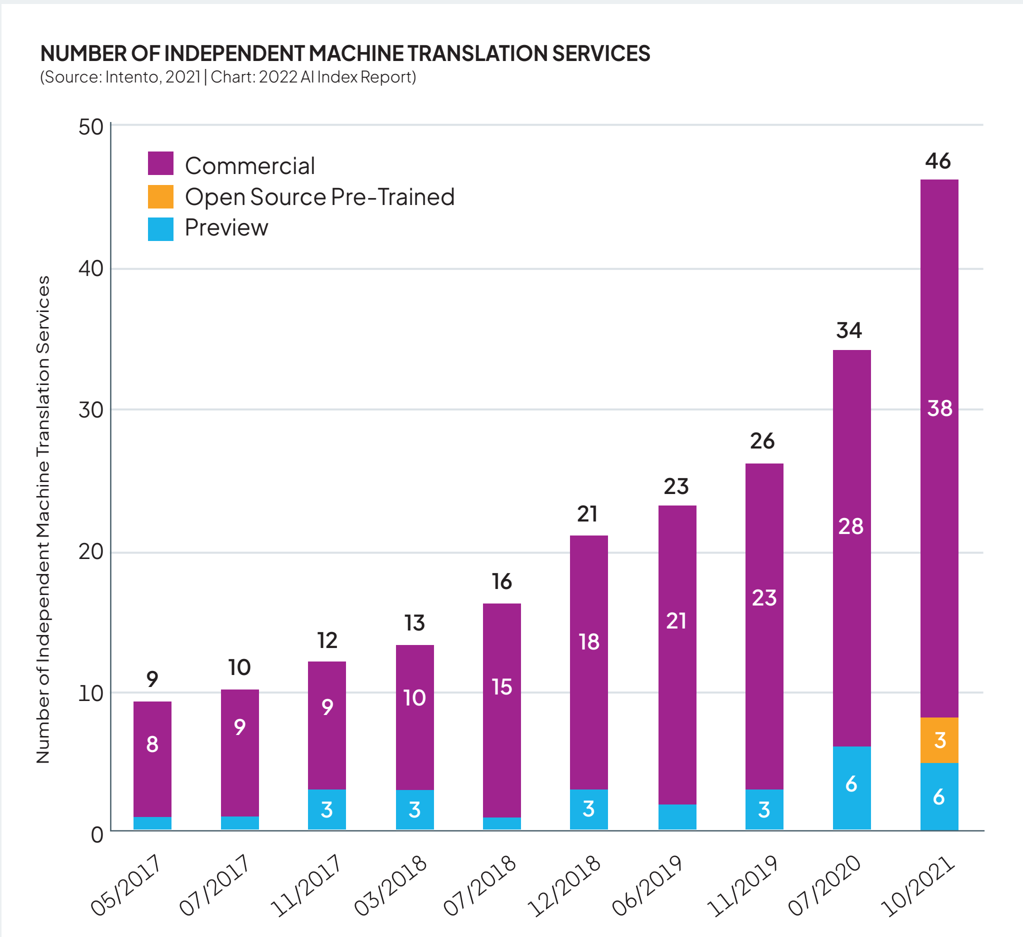
Growth:
Number of Newly Funded AI Companies

Cultural context matters in AI development. The gap between the two leading countries and the rest of the world in AI development is significant – there is a heavy dominance of US based AI development, followed by China and more distantly by select European and Asian countries.



Pre-Trained: Number of Commercially Available MT Systems

The growing interest in machine translation is reflected in the rise of commercial machine translation services such as Google Translate. Since 2017, there has been a nearly fivefold increase in the number of commercial machine translators on the market – but only few pre-trained models. 2021 saw the introduction of the open-source MT services *M2M-100*, *mBART* and *OPUS*.



Rethinking AI As Community Work

Alex Hanna, a sociologist by training, explores how the use of data in new computational technologies is helping to exacerbate existing inequalities around gender, ethnicity and class. We spoke with her about why she left her job on Google's ethics team to join her former supervisor, Timnit Gebru - who had previously been fired from Google.

You have joined the DAIR institute as Director of Research. The institute aims to counter Big Tech's pervasive influence on the research, development and deployment of AI. In what ways is this influence problematic?

Google, Microsoft and Facebook only fund research relating to existing scientific paradigms concerning the optimization of their business models. That's directly or indirectly the case, so either in terms of the types of papers they put out or the funding they give to university researchers, research nonprofits or "AI for Good" projects. Funding guides what problems people work on. They typically don't fund things that are contrary to their interests; and if they do, it's in a very limited capacity.

So, right now, AI is part of the problem?

It is part of the problem when it is used to concentrate and consolidate power and it is used as a means of exacerbating existing inequalities. Most of the time, that AI is implemented in the Big Tech context,the aim tends to be that of facilitating recommendation systems, ad targeting or minimizing customer "churn," so, it's a facilitator for business. AI is also being used in the public sector as a means to minimize the amount of human labor needed for welfare allocation or to identify fraud. But at the same time, it is becoming a tool of surveillance. AI often has the effect of worsening conditions for workers, either by creating a new class of laborers who work for minimal wages to produce data for AI or by optimizing conditions for employers in gig economy settings to the detriment of workers.

What do you plan to do differently at DAIR?

People who are impacted by AI and automated decision-making systems need to have a much greater say in where and when these systems can be deployed. We want to begin by including communities in research activities.

How do you intend to achieve this? How can we start a new discussion on the conscientious use of AI?

If we are going to rethink AI, we will have to rethink what is needed by communities, especially what is needed by marginalized racial, ethnic and gender communities. Some of these tools can be used as a means of taking some of that power back or supporting community decision-making and engagement. Some of the work DAIR is doing points in that direction, for instance, the work that we've done on spatial apartheid and on how AI can support processes of desegregation in South Africa. Another thing that we're looking into is how we can use AI or natural language processing tools to find and identify abusive social media accounts of government actors. We're trying to recalibrate how AI is used and to find a way that doesn't concentrate power but instead redistributes it.

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

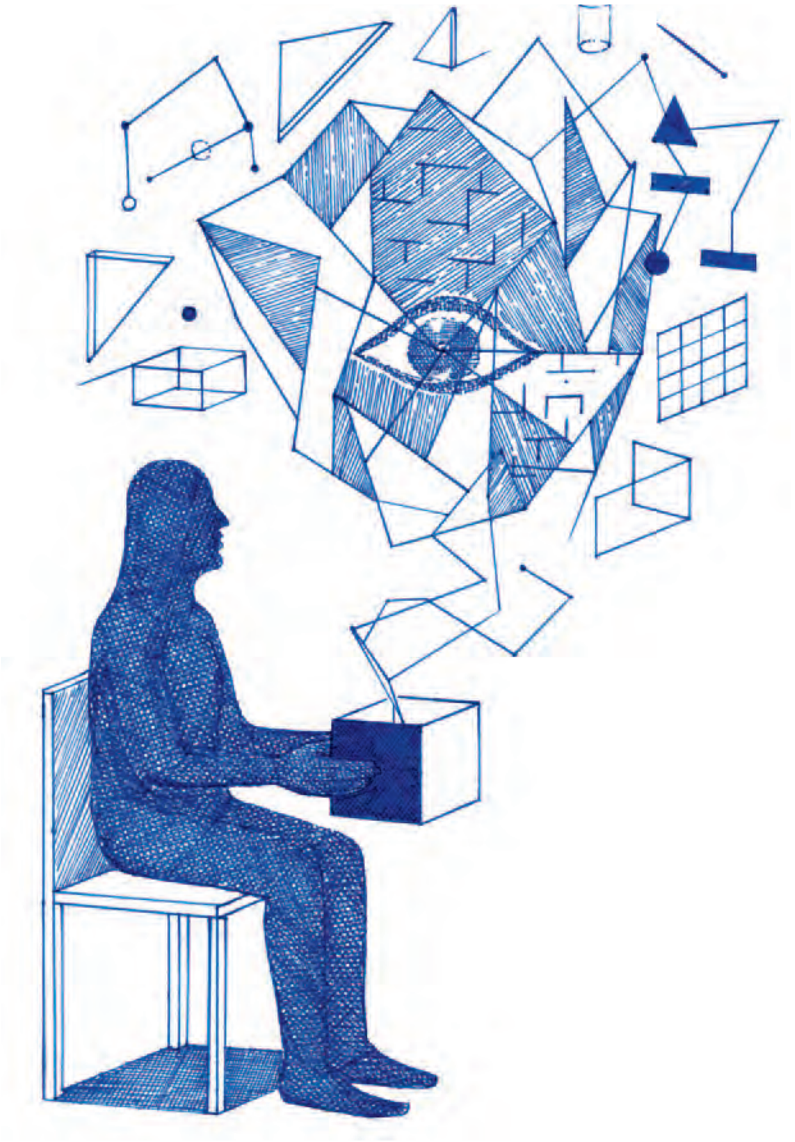
The 2021 paper by Timnit Gebru, Emily Bender, et al., discusses the risks of large language models, AIs trained on a huge amount of text data. Under the right conditions, these models, which are currently quite popular, have become astonishingly good at producing what appears to be meaningful text, and even at extracting meaning from language, or so it seems.

More Harmful than Flying

The "Stochastic Parrots" study builds on previous research work, especially the 2019 paper from Emma Strubell and her collaborators on the carbon emissions and financial costs associated with large language models ("Energy and Policy Considerations for Deep Learning in NLP"). Training large AI models consumes a lot of computer processing power and hence lots of electricity. Their energy consumption and carbon footprints have been exploding since 2017, as models have been fed more and more data. Training a version of Google's language model BERT, which underpins the company's search engine, produced 1,438 pounds of CO_2 emissions, roughly equivalent to a round-trip flight between New York City and San Francisco. Such models aren't only trained once though, but many times over in the research and development process.

Reproducing Social Distortion

Only rich organizations, the paper argues, have access to the resources required to build and sustain such large AI models, while the climate



In December 2021,

on the anniversary of her exit from Google, Timnit Gebru published a press release in which she announced the launch of a new organization, the Distributed AI Research Institute (DAIR), which is designed as "an independent, community-rooted institute set to counter Big Tech's pervasive influence on the research, development and deployment of AI." The institute's work is focused on the process and principles of AI research. One of its premises is that the dangers embedded in AI technology would be preventable if its production and deployment were based on the inclusion of communities and a greater diversity of perspectives.

Today, one of the institute's projects is to use satellite imagery and computer vision to analyze the effects of spatial apartheid in South Africa. In another project, Datasheets for Datasets, Timnit Gebru tries to establish currently non-existent industry standards for documenting Machine Learning datasets. With Datasheets for Datasets, she aims to increase transparency and accountability within the Machine Learning community, mitigate biases in Machine Learning models and help researchers as well as practitioners choose the right dataset.

Continuation …

change effects caused by their energy consumption hits marginalized communities the hardest. The training data is generally collected from the internet, so there's a risk that racist, sexist and otherwise abusive language ends up in it. Because the data sets are so large, it's very difficult to audit them to check for these embedded biases. The authors conclude that large models interpret language in a way that reproduces outdated social norms and patterns of discrimination. These models will also fail to capture the language and the norms of countries and peoples that have less access to the internet and thus a smaller linguistic footprint online.

The Costs of Profit

According to Timnit Gebru and her colleagues, another issue with large language models is the risk of "misdirected research effort." They argue that these models don't actually *understand* language. The models just parrot what was put into them based on the calculated probability that certain words create meaning. They are merely excellent at *manipulating* language. Big Tech companies have continued to invest in them because of the profits they promise. On a social scale, it would be more desirable to work on AI models that might achieve understanding, or that achieve good results with smaller, more carefully curated data sets (and thus consume less energy). But the authors fear that nothing beats the promise of profit, even if large language models come with another risk: They could be used to generate misinformation because they appear to be so meaningful.

Chronicle of a Split Foretold

The developments that led to the foundation of the DAIR Institute aren't just the story of a dismissal. They raised awareness about the existence of a toxic culture in Big Tech.

On the evening of December 2, 2020, Timnit Gebru, the co-lead of Google's Ethical AI team, announced via Twitter that the company had forced her out. She was known for co-authoring a groundbreaking study called "Gender Shades" on the gender and racial biases embedded in commercial face recognition systems in 2018, when she was a researcher at Microsoft. The study showed facial recognition to be less accurate at identifying women and people of color, which means its use could end up discriminating against them. She also cofounded the Black in AI affinity group and champions diversity in the tech industry. Her critical work has frequently challenged mainstream AI practices.

Gebru's departure was the result of a conflict over a paper she co-authored. Google executives asked her to either withdraw a still unpublished paper, or remove the names of all the Google employees from it (five of the six co-authors). Jeff Dean, the head of Google AI, told colleagues in an internal email (which he later shared on Twitter) that the paper "didn't meet our bar for publication" because it "ignored too much relevant research." Specifically, he said it didn't mention more recent work on how to make large language models more energy efficient and mitigate problems of bias. However, the paper's citation list contains 128 references. This contributed to speculations of other actors in the field of AI ethics that Google pushed Timnit Gebru out because the paper revealed some inconvenient truths about a core line of its research. More than 1,400 Google staff members and 1,900 other supporters signed a letter of protest after her dismissal.

The paper in question is called "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?". Emily M. Bender, a professor of computational linguistics at the University of Washington, was the only co-author who was not a Google researcher. The paper's goal, Bender told *MIT Technology Review*, was to take stock of the landscape of current research in natural language processing. Google pioneered much of the foundational research for large language models. Google AI was the first to invent the transformer language model in 2017, which serves as the basis for the company's later model BERT and OpenAI's GPT-2 and GPT-3. BERT now also powers Google search, the company's primary source of money. Bender worries that Google's actions could create "a chilling effect" on future AI ethics research. Many of the top experts in AI ethics work at large tech companies because that is where they find work.

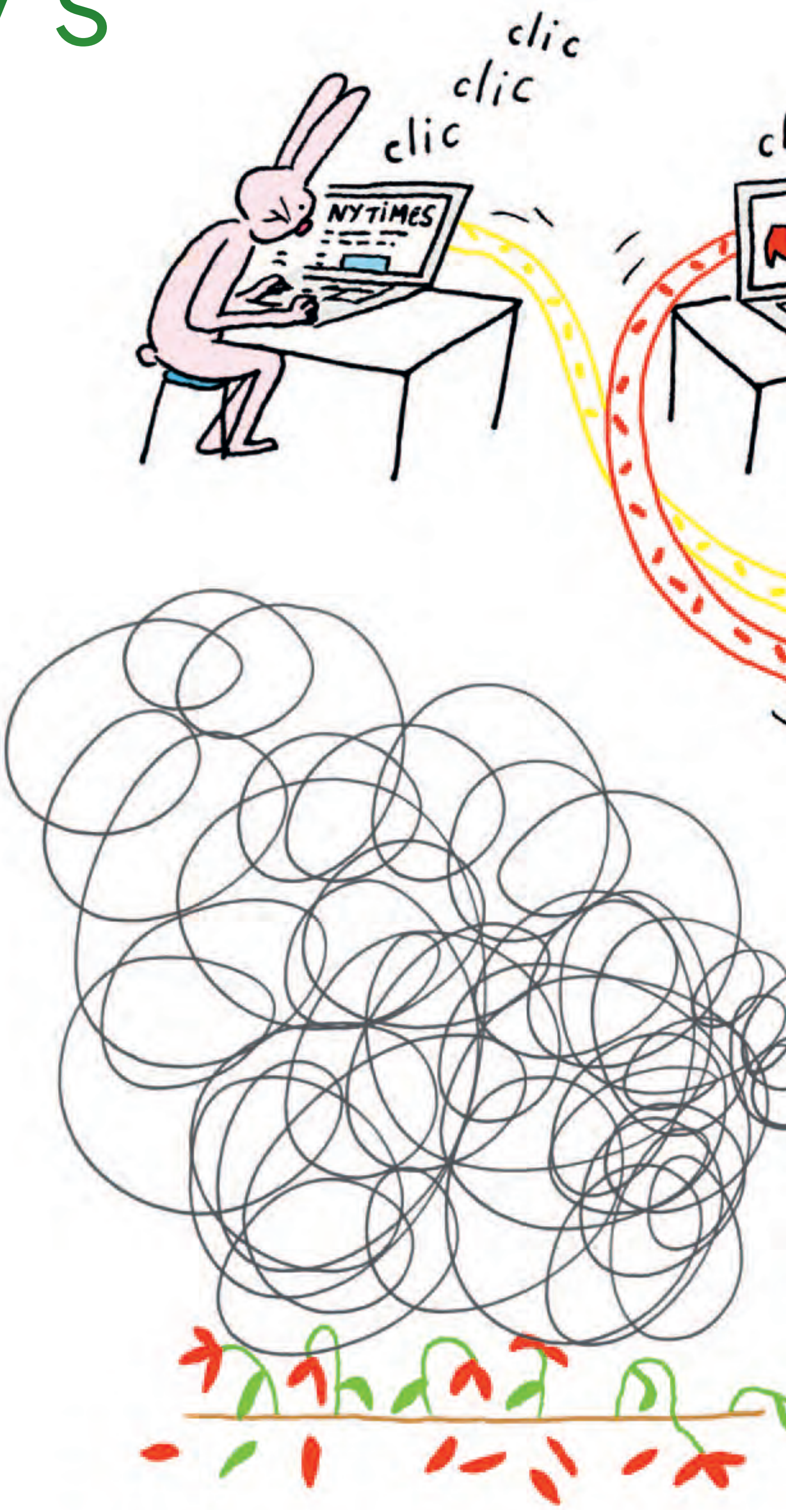
Two members of Google's Ethical AI group have since left Google. Senior researcher Alex Hanna and software engineer Dylan Baker joined Timnit Gebru's nonprofit research institute, Distributed AI Research (DAIR). Hanna announced her resignation on Medium. In her announcement, she criticized the "toxic" work environment at Google and lamented the lack of Black women in the Google Research organization. She concluded: "In a word, tech has a whiteness problem. [...] So in this sign-off, I encourage social scientists, tech critics, and advocates to look at the tech company as a racialized organization. Naming the whiteness of organizational practices can help deconstruct how tech companies are terrible places to work for people of color, but also enable an analysis of how certain pernicious incentives enable them to justify and reconstitute their actions in surveillance capitalist and carceral infrastructures."

DR. ALEX HANNA

... is Director of Research at the Distributed AI Research Institute (DAIR). She has worked extensively on the ethics of AI and on social movements. She serves as a co-chair of Sociologists for Trans Justice, as a Senior Fellow at the Center for Applied Transgender Studies and sits on the advisory board for the Human Rights Data Analysis Group and the Scholars Council for the UCLA Center for Critical Internet Inquiry.

Not a Treat: Cookies Are a Pervasive Technology's Instrument

Carbolytics is a project at the intersection of art and research by artist Joana Moll in collaboration with the *Barcelona Supercomputing Center* (BSC). The project aims to raise awareness of the environmental impact of pervasive surveillance within the advertising technology ecosystem (AdTech).



Why did you choose to investigate the carbon footprint of AdTech?

AdTech is the primary business model of the internet and the carbon costs associated with it are highly opaque. Companies need to be held accountable for what they're doing. The AdTech industry is just the tip of the iceberg. There is a massive ecosystem beyond cookies and beyond browsers. We have no clue how user data are being exploited and how big the energy consumption of such a huge business model is.

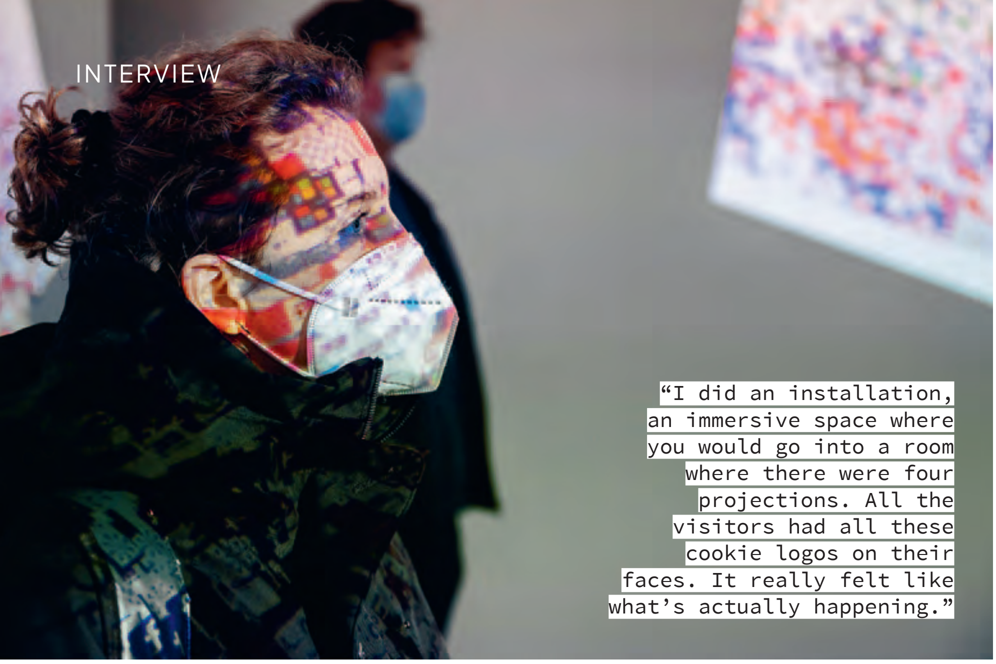
What did you learn through the project?

In most cases, it was very difficult to identify the organizations behind the cookies. And we found that the most pervasive cookies like Google Analytics are not the most polluting. We also found out that the most polluting website when it comes to cookies was Netflix, but there weren't many of them.

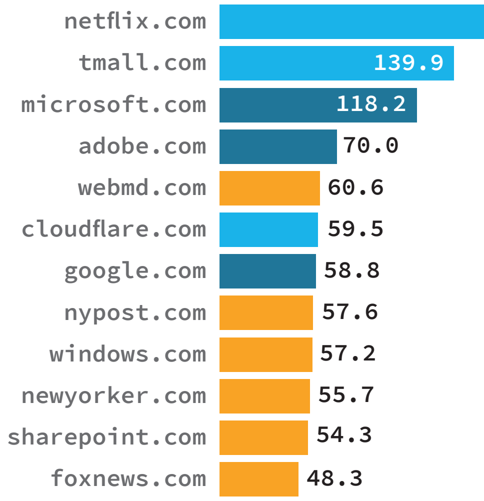
»Most of our transactions are being quantified and commodified. AdTech is taking advantage of this by exploiting everything we do online. Every mouse movement, every word we type, every click is capitalized on and ultimately generates revenue.«

How is it possible that such a carbon emission-intensive technology like AdTech is an ecological blind spot?

It's an extremely fast process. Everything happens extremely fast and it's overshadowed by all the devices that we use. If we were aware of every single process that happens, the system probably wouldn't work. The problem is that we don't really understand how things work. In Slovenia, I did an installation, an immersive space where you would go into a room where there were four projections. All the visitors had all these cookie logos on their faces. It really felt like what's actually happening. A lof of people were overwhelmed by the fact that nobody really knows what to do to counter this. It's frustrating because we as individual users can't do anything about it. It's a systemic problem. It's not only very hard to understand what the cookies do, it's a crazy amount that we are exposed to.



“I did an installation, an immersive space where you would go into a room where there were four projections. All the visitors had all these cookie logos on their faces. It really felt like what’s actually happening.”



You see AdTech as a part of cognitive capitalism. What does this mean?

Cognitive capitalism is the economic system that we are all part of. In this system, wealth is no longer being produced exclusively by material goods but through intangible actions, experiences, communication and cognition. Most of our transactions are being quantified and commodified. AdTech takes advantage of this by exploiting everything we do online. Every mouse movement, every word we type, every click is capitalized on and ultimately generates revenue.

Did you find anything you didn’t expect?

Consent cookies, the ones that ask for user consent to take data, were the third most pervasive we found. It is quite perverse that privacy adds an extra layer to the issue. This is why I think that privacy and sustainability always need to be conceived together. They are part of the same problem: the lack of accountability of polluting companies.

What do you think of announcements such as the one made by Google, that the company will be carbon positive by 2030?

I think it’s really problematic that they can legally say this, because it’s impossible. Google is not just the operations data center. Google is on all our devices. So, there’s no way they can quantify all their energy consumption. How would it be possible that all this technology even reaches carbon neutrality? It’s not possible because they feed from our devices. When we calculated the carbon emissions of cookies, we found that there is not enough independent assessment of energy consumption and carbon emission of data in general. And researchers dramatically disagree on how to quantify this, which I believe is a huge problem.

The interactive web-based installation Carbolytics shows the average global volume of cookie traffic in real time and the energy consumption associated with it. We see how cookies parasitize user devices to extract personal data. (https://carbolytics.org/web2x)

428.6

Top sites per CO_2 emissions
[tons per month]

What is your main takeaway from Carbolytics?

Something interesting happened with this project. I was expecting much more interest from the media because it's about the primary business model of the internet. I talked to very big newspapers, but nobody would pick up or follow up on the story, which was very frustrating. Then I saw that *The New York Times* and similar companies had a lot of cookies that appear within the top 20 in our ranking. I was pretty sure that in the end, I was ignored because the story touched their own business model. It's difficult to raise public awareness if the media – which is supposed to be the gateway for reaching the public – isn't interested in spreading the news because the news affects them so much.

JOANA MOLL
The artist behind Carbolytics

About Carbolytics
(Excerpt from the project description by Fernando Cucchietti, Joana Moll, Marta Esteban, Patricio Reyes, Carlos García Calatrava)

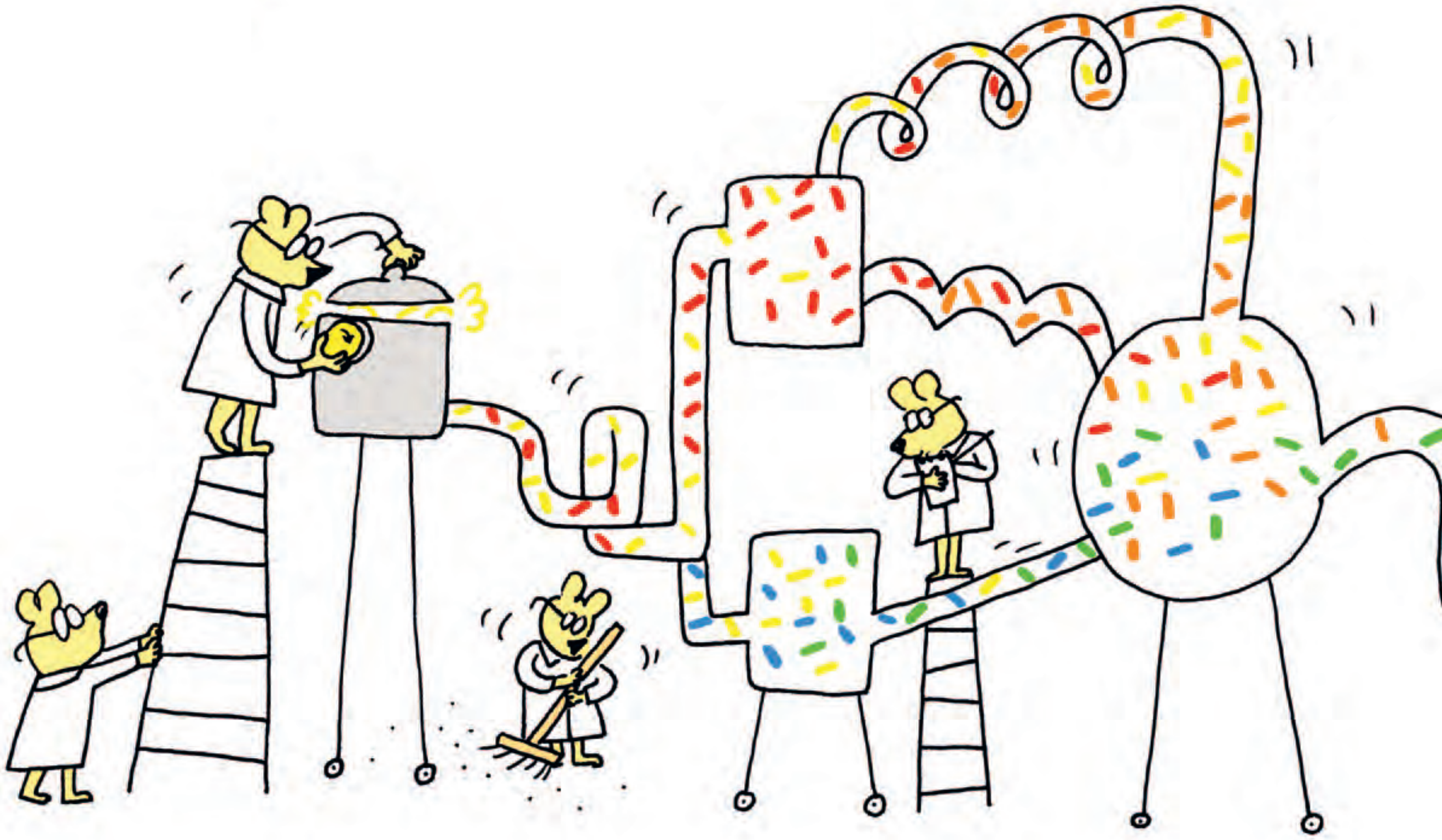
Tracking users' online behavior has become a major business model in the last decade. Online tracking is the act of collecting data from online users as they read the news, purchase items, interact on social media, or simply perform online searches. Companies rarely disclose information on the environmental footprint of such operations. This expansive data collection often becomes the basis on which AI operates.

AdTech analyzes, manages and distributes online advertising and is the primary business model of the data economy ecosystem. In 2021, global ad spending across platforms reached $763.2 billion, and it is expected to rise 10 percent in 2022. In 2020, 97.9 percent of Facebook's and 80 percent of Google's global revenue were generated from advertising, and these companies, together with Amazon, will dominate 80 to 90 percent of the market in 2022, excluding China. Yet, despite the extraordinary relevance of AdTech within the global economy, its methods and processes are extremely opaque, and thus difficult to control and regulate. Data collection through AdTech often becomes the prerequisite for AI applications, such as recommender systems, and thus needs to be accounted for when considering the sustainability of AI.

Typically, data is collected through cookies and other tracking technologies integrated into devices, web pages, apps and all kinds of interactive and audiovisual digital content. Even though they are created and stored on the user's device, tracking technologies are mostly non-transparent and sometimes undetectable to users. Despite their "invisibility" and relatively small size, tracking technologies are responsible for triggering myriad algorithmic processes at a global scale, exploiting user behavior data with a direct impact on the user's devices power consumption.

The research behind Carbolytics analyzes the carbon emissions of all those cookies belonging to the top 1 million websites. The investigation identified more than 21 million cookies per single visit to all these websites, belonging to more than 1,200 different companies, which translates to an average of 197 trillion cookies per month, resulting in 11,442 monthly metric tons of CO_2 emissions. This number reflects only browser-based cookie traffic and does not include other behavioral advertising tools, such as app tracking activity or profiling algorithms.

»Complex data processing requires more time, staff, and expensive systems.«



Less Is More: Why a Data Diet Can Benefit Artificial Intelligence

Computational linguist Michaela Regneri has investigated AI e-commerce applications for the Hamburg-based online shopping platform OTTO. She has developed recommendation algorithms that are strategically relevant for the shopping platform while limiting the need for resources in AI development. Regneri is an advocate of data minimalism, and she wants to establish it as an efficiency strategy for Artificial Intelligence in the data economy. Her aim is to use minimal data while at the same time preventing a drop in data quality.

What is data minimalism and what are the advantages of working in a data minimalist way?

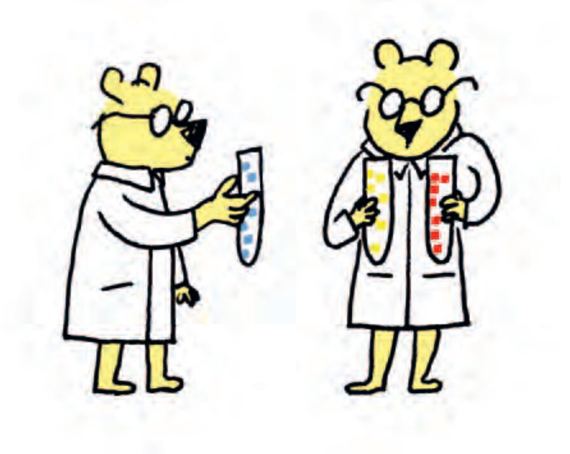
Minimalism refers to the amount of data we process with AI. We started with the question: What is the actual value of the data we process? As we aim for efficiency, that question is interesting both economically and ecologically. We want good, reliable AI that does what it is supposed to do and is as efficient and effective as possible. So, the less data needed for the same procedure, the better the result.

In our year-long project for OTTO, we first examined how we could determine which of the various data points or how much of the entire data collection was useful for the AI. Our aim was to determine the use-based value of the data. We conducted practical experiments and tested the individual algorithms to ascertain if feeding them with data really helped the system. Which data lead to better results for the algorithm and which make it worse? We saw that, beyond a certain point, saturation occurred, and adding new data became unproductive.

How do you slim down the amount of data?

We tried different ways. The simplest method is to leave out data and see how well the algorithm works with half the data initially, then with one-third of the data or with two-thirds of the data. This is a sensitivity analysis, and it works in principle like a reverse allergy test. With allergies, we look for harmful influences; we identify data that have a positive influence. If a data point is omitted and the system improves, then it was probably a bad data point. If it is omitted and nothing happens, then we know it is unnecessary for the system to function. And if the system gets worse, we have discovered a valuable data point.

There is a widespread tendency to throw in all the data you can get – particularly in industrial environments, where there is a lot of data, sometimes way more than you need. Sure, you might achieve an incremental improvement of half a percent, but to make the algorithm really efficient, you need to remove the harmful data. For example, we have had the problem of bots being let loose on a webpage. They wildly click around the shop, causing specific items to



»We select those data that actually contain important information, and make sure that the group is not larger than necessary for the desired result.«

suddenly look very popular. This is bad for the system, of course. Sometimes, we unintentionally harm the AI ourselves, for example with a marketing promotion like a deal of the day, which results in lots of people clicking on a cheap article, when actually they are not necessarily interested in the item itself, just the discount. So, we are not dealing with their natural buying behavior, but with their reaction to something we did. This makes it difficult to draw any useful conclusions about the AI's behavior.

What are the economic advantages of data minimalism?

Data minimalism reduces costs. Complex data processing requires more time, more staff and expensive systems. Data protection and compliance add to the cost. The less data that is managed in the cloud, the lower the cost of the cloud. Because the innovation loops become shorter, the model can be trained faster and has a greater innovation potential. The algorithms can be tested faster, and ultimately, eliminating harmful data points leads to better performance. In an economic context, this always means more profit.

Dimension: environmental sustainability

Criteria: energy consumption

Indicator: measures are used to reduce the amount of data

The energy consumption of an AI system differs depending on the particular phase in its life cycle. The development phase of new AI models can be extremely energy intensive, despite growing hardware efficiency. Finding the desired model architecture, in particular, can sometimes require tremendous computing power. Energy consumption in the training and, especially, the application phase is significantly lower. However, unlike development, which is just a one-time process, these phases are sometimes repeated a massive number of times. Data minimalist approaches, which keep the data sets used for training and application small, are one way of reducing energy requirements during the training and application phases.

Are there also social benefits to data minimalism?

Reducing data means improving privacy and cybersecurity. Every sensitive data point that gets moved around unnecessarily is an extra security risk. Moreover, data minimalism can help curb discriminatory patterns in data sets.

How so?

People often argue that you need a lot of data for data sets to be balanced and non-discriminatory. There are, however, a few older techniques, based on making data sets smaller, that avoid discrimination. These come from medicine and medical statistics. For medicine, the most interesting group is the smallest one: the people who are actually sick. In order to do justice to this group in medical applications, it is possible to make the larger group of healthy people smaller. This can be understood from a data minimalist perspective: We select those data that actually contain important information, and make sure that the group is not larger than necessary for the desired result. In this way, minorities or smaller groups are given more weight, in relative terms.

Rigorous data minimalism would result in AI systems that allow us to account precisely for the effects that individual data points or data sets have on the AI system. This in turn means that we could also predict which data have a discriminating influence on the result.

In retrospect, how would you assess your experience at OTTO? What insights did you gain with regard to the development possibilities of sustainable AI?

For me, the exceptional aspect of this project was the ability to do applied research. Universities have limited data available. For large corporations, there is no problem managing a lot of data, because they have the computing and financial resources. At OTTO, we were able to use this data to build AI algorithms ourselves. But we didn't want to blindly stuff everything into the system just because we could. Less data means less computing time and lower CO_2 emissions, simply because consumption is being reduced. Energy consumption runs through the entire value-added chain, and power consumption grows linearly with the amount of data processed in the algorithm.

However, there are no reliable and proven methods, either in science or in industry, that take these consequences into account and minimize them, and we need to do something about that. If we want AI and sustainability to become a reality, government agencies need to promote the kind of collaboration that we had at OTTO between science and industry.

Dimension:
social sustainability

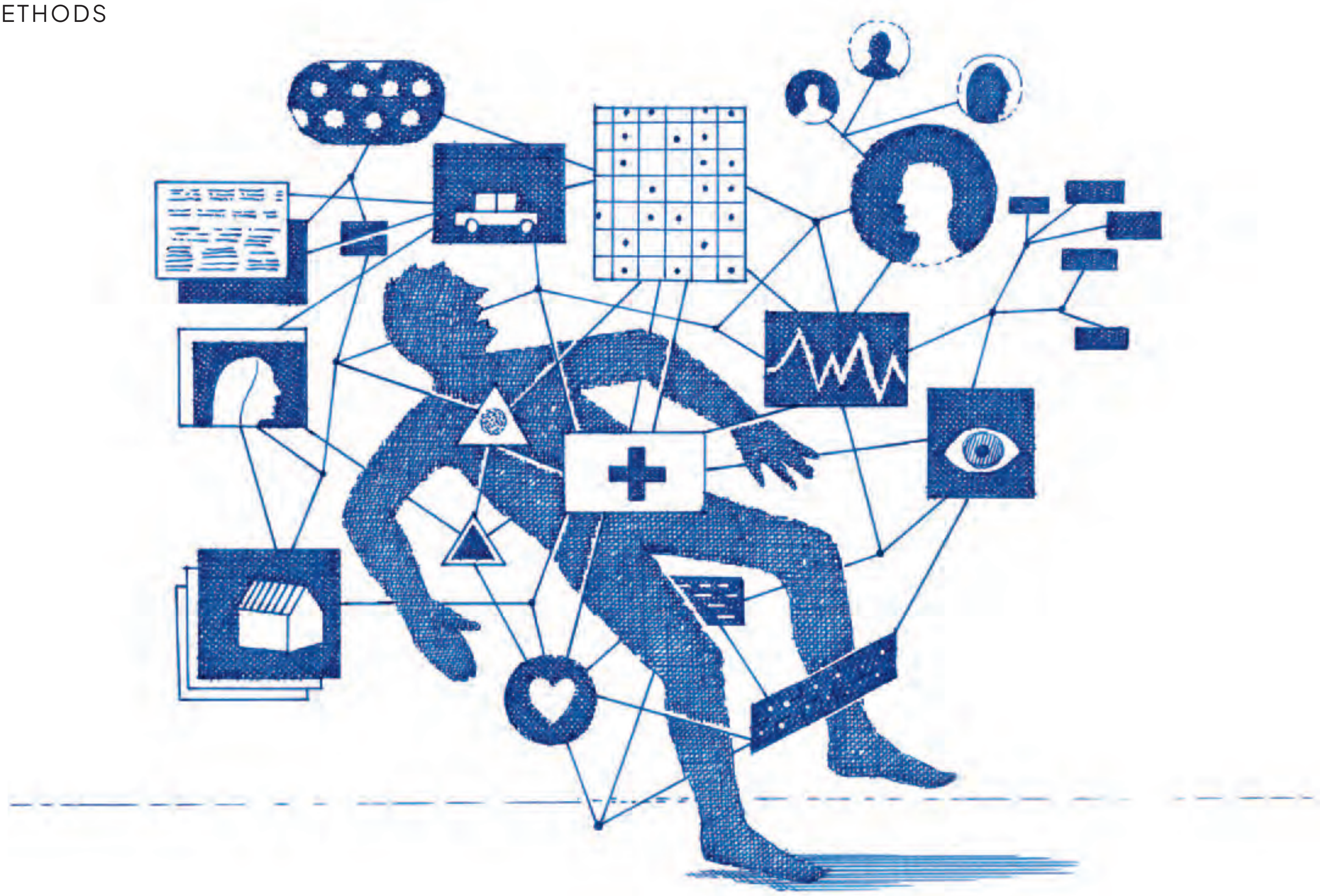
Criteria:
self-determination and data protection

Indicator:
the consistent implementation of a Privacy by Design approach

A consistent Privacy by Design approach takes data privacy and security interests into account during the planning and development stage of digital technologies. The General Data Protection Regulation (GDPR) specifically calls for a Privacy by Design approach, although it does leave room for flexibility in implementation. Privacy by Design means, for example, that data is encrypted and anonymized, used sparingly and not merged unnecessarily with other data. Data minimalism is thus an integral part of Privacy by Design. With data-minimalist AI development, good data management and high-quality selected data discrimination risks in AI applications can be reduced. This also saves resources when it comes to the computing power necessary.

DR. MICHAELA REGNERI

… is passionate about Artificial Intelligence. Her main interests lie in the fields of cognitive computing with language, images and all other types of human-machine interactions. After completing her doctorate in computational linguistics, she was initially responsible for search and data mining at the SPIEGEL-Verlag publishing company. In 2016, she began working for OTTO as a Product Manager for business intelligence analytics. Michaela Regneri is particularly interested in topics related to corporate digital responsibility and organizational development, including AI and the future of work, AI and staff development and the sustainable design of AI systems.



Automatic Oblivion

Responsible Data Management in Machine Learning

Having a say over one's own data and being able to exercise one's right to be forgotten is essential, especially regarding the increasingly pervasive use of AI systems in our everyday life. As a recent case reported in the Netherlands demonstrates, consequences can be devastating if this right isn't enforced on time. A couple lost their baby through a miscarriage and were subsequently continuously exposed to baby product ads online – a traumatizing experience. They contacted the e-commerce company that was advertising the baby products, but were told that the company had no technical means of disabling the online recommendations.

Such online recommendations are based on automated decision-making (ADM) systems – AI applications increasingly used to automate decisions that have an impact on people's lives. These systems are based on sample data from which prediction models are derived with Machine Learning techniques. Such ADM systems are used in domains as varied as credit and lending, medical diagnosis and hiring. The risks and opportunities linked to their widespread use are garnering much attention from policy-makers, scientists and the media. An important subset of these risks arise from technical challenges with respect to the management of the data stored and processed by ADM systems. As the example above shows – privacy protection and the ability to have a say over one's personal data is essential.

TECHNICAL BIAS INTRODUCED BY COMPUTATIONAL SYSTEMS

Much of the current discussion about algorithmic fairness of automated decisions focuses on so-called pre-existing bias, which has its origins in society. In ADM systems, this type of bias often exhibits itself in the sample data for prediction models. Technical bias arises in the data due to the technical system's operations. The risks of introducing technical bias in data-driven ADM abound, but a technical fix is possible – which isn't the case with pre-existing bias.

Dimension:
social sustainability

Criteria:
self-determination and data protection

Indicator:
ensuring informational self-determination

People must be enabled to maintain autonomy over their personal data. This can work through simple visualizations, notifications, and consent or revocation mechanisms. Users should be informed as soon as AI systems use or collect personal data. They should have a say in the use of their data and not be restricted in their self-determined actions by mechanisms that influence behavior, such as nudging or dark patterns.

Technical bias is a consequence of the "lab conditions" under which experts usually design the algorithmic component of an ADM system. They work with a fixed and clean dataset of training examples and try different algorithmic approaches to find a prediction model that works well on this dataset. However, once the full ADM system is developed and applied in the real world under "production conditions," the way in which data is produced for the prediction model changes.

ADM systems typically process data from multiple sources (oftentimes other technical systems) that continuously produce new data. The prediction model must regularly be adjusted to the new data. The system must combine the data from all sources and prepare it in a form that the prediction model understands. In this data preparation process, technical bias can be introduced by programming errors or a misrepresentation of groups in the generated data. It can even be introduced through seemingly innocuous operations, for instance if demographic data is filtered by zip code during data preparation, as the place of residence potentially correlates with sensitive demographic factors such as age, income level or ethnicity. As a result, the prediction model might produce less reliable predictions for groups of individuals that are not well represented in the data due to filtering operations.

ENFORCING THE "RIGHT TO BE FORGOTTEN" IN A TIMELY MANNER

An orthogonal dimension related to data management in ADM systems is their compliance with laws guaranteeing privacy and digital self-determination rights. A prominent example is the "right to be forgotten" (Article 17 of the General Data Protection Regulation, GDPR). It requires companies and institutions that process personal data to delete user data upon request: "The data subject shall have the right to [...] the erasure of personal data concerning him or her without undue delay [...] where the data subject withdraws consent." The GDPR law does not specify how soon data must be erased after a deletion request, yet it states the "obligation to erase personal data without undue delay" using "appropriate and effective measures." Currently, data erasure seems to be a rather tedious and lengthy process in practice. The erasure from active systems in the Google cloud, for example, can take up to two months.

This is why we need ADM systems with "unlearning" capabilities enabling them to delete users' interaction data on request and thereby to adjust their predictions on demand. But this poses many challenges with respect to the algorithmic and computational efficiency of "updating" existing prediction models.

Academia and industry have just started addressing these challenges (in part motivated by pending regulations). But even if technological foundations for a responsible data management in ADM systems were established, we would still need best practice solutions. In order to find them, it is crucial to have access to real-world ADM systems, which we don't presently have because most of these systems are proprietary and run by private enterprises.

DR. SEBASTIAN SCHELTER

… is an Assistant Professor at the University of Amsterdam, conducting research at the intersection of data management and Machine Learning. In his work, he addresses data-related problems that occur in the real-world application of Machine Learning. Examples are the automation of data quality validation, the inspection of Machine Learning pipelines via code instrumentation or the design of Machine Learning applications that can efficiently forget data. Schelter makes most of the research code that he writes available under an open source license and is an elected member of the Apache Software Foundation.

We Are(n't) AI (Yet)

Julia Stoyanovich has dedicated her career as a computer scientist to responsible data management and responsible AI. She is a fierce advocate for educating the public about the impact that AI and algorithms have on their lives – by offering free public library courses and even creating comics.

»With the degree of automation that is already present in that industry, every one of us has either already been affected or will be affected by AI systems when looking for a job.«

You are the founder of the Center for Responsible AI. The Center's claim is "We build the future in which responsible AI is the only AI." How close are we to that goal?

Is responsible AI synonymous with AI today? Unfortunately, not yet. We are still quite far from that. It will take a combination of approaches and solutions to get there, and some of the solutions will be technical. Better algorithms built by people who are aware of the issues will help. But I think at this point, the main gap is not really algorithmic.

What else is needed?

We don't have a shared understanding of the role AI should play in society. And in order to reach that understanding, we need to get people to understand what AI can and cannot do. Among the public at large, there is a lot of magical thinking about AI. This hands a lot of power to the people who are developing this technology - perhaps even power they don't really want, because they are not necessarily trained in law or the social sciences. We shouldn't put technologists in a position where they have to adjudicate some of these issues with AI through code.

Do we need a broader discussion on what AI can do for us?

We are running a course right now called "We Are AI." It's a public education course that we offer through the Queen's Public Library in New York City, available and accessible for everyone, irrespective of their background, math knowledge or programming experience. The course is free of charge; we even give people gift cards for attending. The message is: "AI is what we want it to be." My colleagues with whom I developed this course, as well as the students, say: "But this is not the case right now. We don't have control over AI." Indeed, today we are, for the most part, subjected to these systems. But by making the claim that "AI is what we want it to be," we assert ourselves. We should really be speaking in the present tense and very forcefully about what we want the world to look like. And by doing that, we begin moving in the right direction.

The Center for Responsible AI is all about transparency. How can we achieve meaningful transparency and governance of AI systems?

Nutritional labels as inspiration for transparency on ranking algorithms

Automated decision-making (ADM) systems often calculate scores and rankings to present their results. For example, a person's creditworthiness is often assessed through the automated determination of a score, or applicants for a job are listed by a ranking algorithm according to their supposed qualifications and suitability for a position. Such scores and rankings are known to be unfair, easy to manipulate and not very transparent. Moreover, they are often used in situations for which they were not originally designed – which can lead to inaccurate and problematic results. For this reason, Julia Stoyanovich and colleagues have developed a rating system that provides information on ranking algorithms similar to nutritional labels for food. The Ranking Facts application uses visualizations to create transparency. It shows things such as the decision-making criteria included in a ranking, how they are weighted and how stable and fair the calculations are. The application is intended to also help non-experts evaluate the quality and suitability of a ranking.

Source:
https://dl.acm.org/doi/10.1145/3183713.3193568

Dimension:
social sustainability

Criteria:
transparency and accountability

Indicator:
publicly available information about system implementation

When people have become the subject of an automated decision, they must be informed of that fact. Similarly, relevant information about AI systems must be made publicly available so that the functionality, the decision-making criteria and technical dependability of the system can be verified by independent bodies. The minimum standard is to document the most relevant information regarding the system's goals, user and usage cases, training and test data, model used, feature-selection processes, inputs, tests, metrics and so on. Such information can be stored in public registers.

One of the constituencies that I've been thinking about when developing transparency and interpretability mechanisms are regular members of the public. I'll give you an example: There are lots of algorithmic systems being used at various stages of the hiring process. We all look for jobs at some point in our lives. With the degree of automation that is already present in that industry, every one of us has either already been affected or will be affected by AI systems when looking for a job. Access to jobs is access to an essential economic opportunity, it's not optional. This is a crucial domain in which we should be figuring out how to regulate the use of AI. We need to find ways to educate job seekers about the systems that they are subjected to, and to explain to employers – companies that are buying such tools – what they are paying for. Regarding the employment sector, we have to know what the algorithms do exactly. Do they look for candidates who may or may not know that they're even a good match? Do they match resumes to positions? Do they determine the level of compensation? All of these possibilities have different requirements and different error margins, and different tools are used in all of them. So you have to be very specific about the domain while speaking to the stakeholders. This is key.

What might this look like in practice?

I've been developing "nutritional labels" for job seekers, to enable different kinds of interaction between a job seeker and the hiring system. One specific label would accompany a job ad. I call it the "posting label". These labels should document what the position's qualifications are so that you can tell whether you're qualified, what data the potential employer will use to screen you, what screening methods they will use, whether you can opt out at any point and what the features are that a particular screener will look at. This kind of label is going to enable informed consent on the part of the job seeker, helping them decide whether to apply for the job, and to request accommodations or an alternative form of screening if necessary. Another kind of label, called the "decision label," would accompany the decision about whether you are hired or denied a job. These labels will tell the job seeker what they can do to improve their chances of being hired in the future. And they will support recourse, allowing the job seeker to contest or correct the decision if, for example, incorrect data was used.

Do you see room for more regulatory approaches in that regard?

One of the things the Center for Responsible AI is very proud of, and I am personally very proud of, is that we were big proponents of a law that was passed in New York City on December 21, 2021, that sets the first precedent for regulation in the domain of algorithmic hiring. The law requires that automated decision-making tools used for hiring and employment be audited for bias. And it also requires that applicants be notified before they apply that such a tool will be used for screening, and also what features of their application it will use. The law will come into force in January 2023, giving the vendors of such AI systems and their users a year to figure out how to comply. I see this law as a great first step.

PROF. JULIA STOYANOVICH

... is an Associate Professor of Computer Science & Engineering and of Data Science, and and she is Director of the Center for Responsible AI at New York University. Her research focuses on operationalizing fairness, diversity, transparency and data protection in all data lifecycle stages. She is active in AI policy, having served on the New York City Automated Decision Systems Task Force and contributed to New York City regulation of AI systems used in hiring. Stoyanovich teaches responsible AI to data scientists, policy-makers and the general public, and is a co-author of an award-winning comic on this topic. She is a recipient of an NSF CAREER award and a senior member of the ACM.

Dimension:
social sustainability

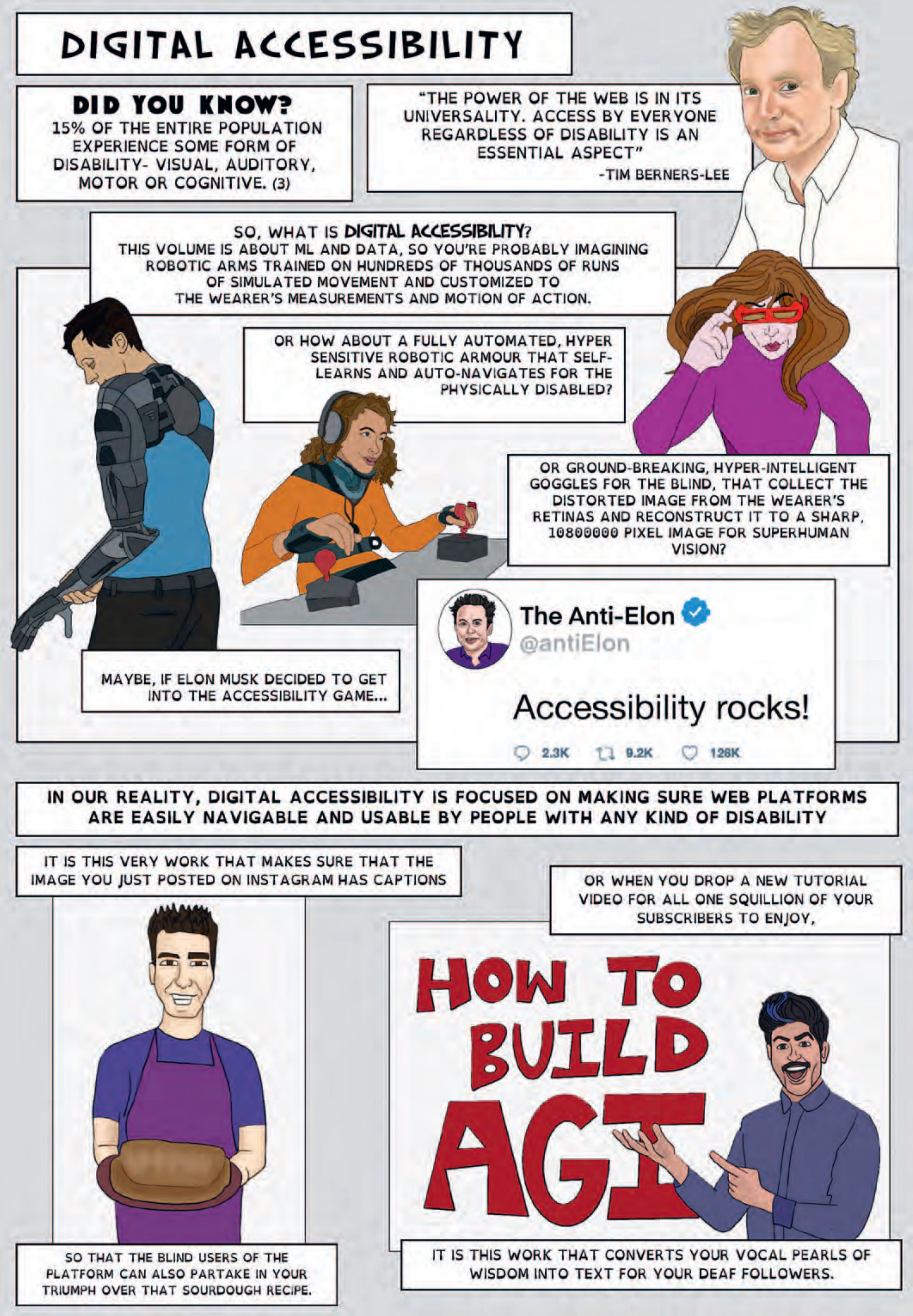
Criteria:
non-discrimination and fairness

Indicators:
detection of, awareness and sensitization to fairness and bias

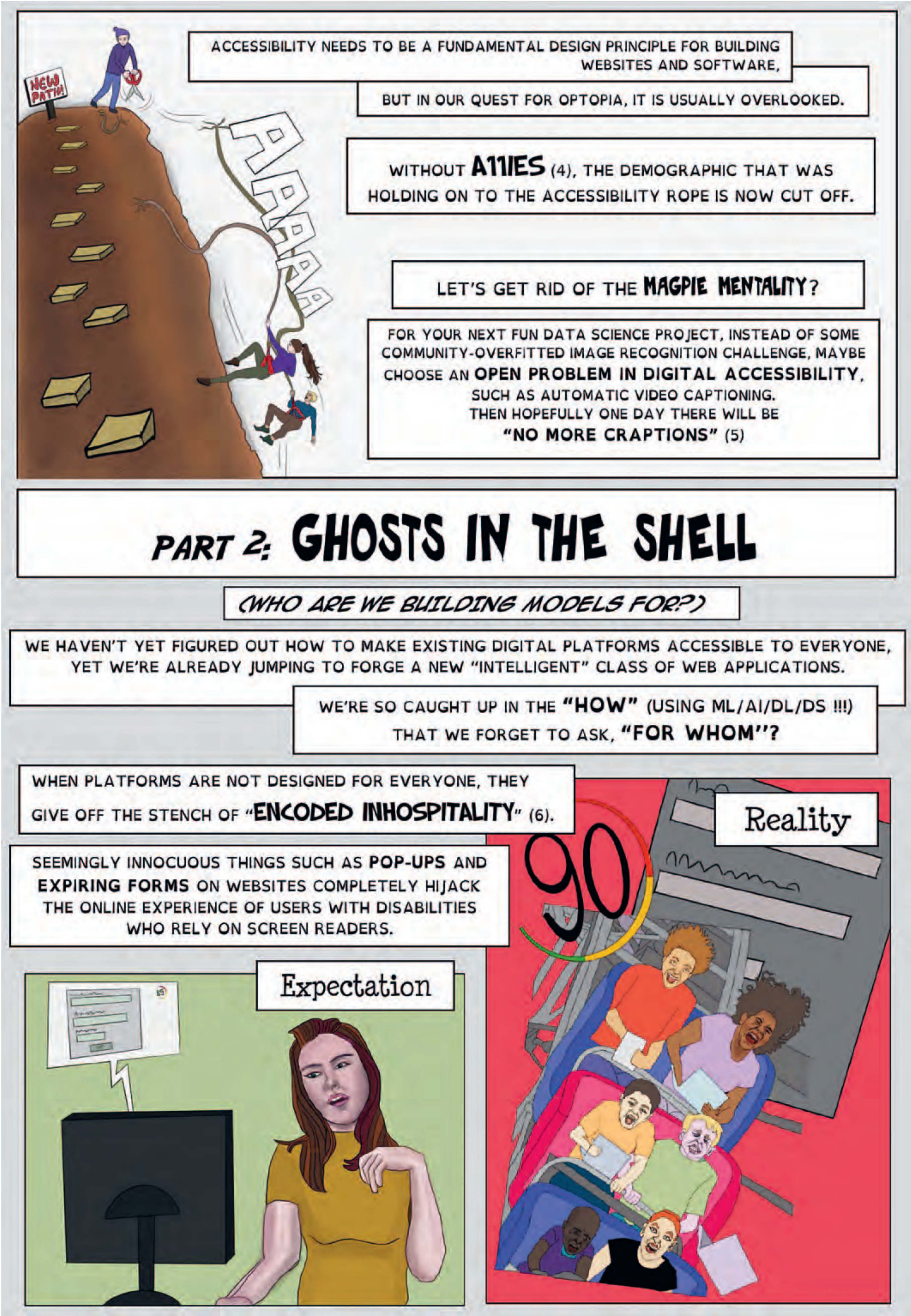
To establish non-discrimination and fairness in the context of AI, organizations that develop or use AI need to raise awareness. The first step is to define fairness on a case-specific basis, and to communicate this definition broadly in the planning and development process. Potential discrimination can be recognized during the development phase of AI systems through impact assessments. There are proven methods for measuring fairness and bias, such as Equalized Odds and Equal Opportunities. These can be used to identify biases in training and input data, as well as in the models, methods and designs, and to make improvements during development. Fairness tests must take into account protected attributes such as ethnicity, skin color, origin, religion, gender, etc. in order to prevent discrimination on the basis of these attributes. The same applies to so-called proxy variables that correlate with the protected attributes.

Page 36 and 37: "Mirror, Mirror" by Falaah Arif Khan and Julia Stoyanovich. Data, Responsibly Volume 1 (2020)
Julia Stoyanovich has published two comic book series in collaboration with Falaah Arif Khan, who is both an artist and a data scientist. One series is called "We Are AI." It was written in English, has recently been translated into Spanish and will appear in other languages as well. This series accompanies a public education course targeting an adult audience without any background in technology, but with an interest in AI and its social impacts. By teaching through comics, a less standard medium that allows for humor, Stoyanovich and Arif Khan try to make AI and technology ethics more accessible. The second comic book series, called "Data, Responsibly," is aimed at data science students or enthusiasts and is slightly more technical. Stoyanovich uses this series as part of assigned reading for the responsible data science courses she teaches to graduate and undergraduate students at New York University.

Download:
https://dataresponsibly.github.io/comics/vol1/mirror_en.pdf



DIGITAL ACCESSIBILITY
DID YOU KNOW? 15% OF THE ENTIRE POPULATION EXPERIENCE SOME FORM OF DISABILITY- VISUAL, AUDITORY, MOTOR OR COGNITIVE. (3)
"THE POWER OF THE WEB IS IN ITS UNIVERSALITY. ACCESS BY EVERYONE REGARDLESS OF DISABILITY IS AN ESSENTIAL ASPECT" -TIM BERNERS-LEE
SO, WHAT IS DIGITAL ACCESSIBILITY? THIS VOLUME IS ABOUT ML AND DATA, SO YOU'RE PROBABLY IMAGINING ROBOTIC ARMS TRAINED ON HUNDREDS OF THOUSANDS OF RUNS OF SIMULATED MOVEMENT AND CUSTOMIZED TO THE WEARER'S MEASUREMENTS AND MOTION OF ACTION.
OR HOW ABOUT A FULLY AUTOMATED, HYPER SENSITIVE ROBOTIC ARMOUR THAT SELF-LEARNS AND AUTO-NAVIGATES FOR THE PHYSICALLY DISABLED?
OR GROUND-BREAKING, HYPER-INTELLIGENT GOGGLES FOR THE BLIND, THAT COLLECT THE DISTORTED IMAGE FROM THE WEARER'S RETINAS AND RECONSTRUCT IT TO A SHARP, 10800000 PIXEL IMAGE FOR SUPERHUMAN VISION?
MAYBE, IF ELON MUSK DECIDED TO GET INTO THE ACCESSIBILITY GAME...
The Anti-Elon
@antiElon
Accessibility rocks!
2.3K 9.2K 126K
IN OUR REALITY, DIGITAL ACCESSIBILITY IS FOCUSED ON MAKING SURE WEB PLATFORMS ARE EASILY NAVIGABLE AND USABLE BY PEOPLE WITH ANY KIND OF DISABILITY
IT IS THIS VERY WORK THAT MAKES SURE THAT THE IMAGE YOU JUST POSTED ON INSTAGRAM HAS CAPTIONS
SO THAT THE BLIND USERS OF THE PLATFORM CAN ALSO PARTAKE IN YOUR TRIUMPH OVER THAT SOURDOUGH RECIPE.
OR WHEN YOU DROP A NEW TUTORIAL VIDEO FOR ALL ONE SQUILLION OF YOUR SUBSCRIBERS TO ENJOY,
HOW TO BUILD AGI
IT IS THIS WORK THAT CONVERTS YOUR VOCAL PEARLS OF WISDOM INTO TEXT FOR YOUR DEAF FOLLOWERS.



ACCESSIBILITY NEEDS TO BE A FUNDAMENTAL DESIGN PRINCIPLE FOR BUILDING WEBSITES AND SOFTWARE,
BUT IN OUR QUEST FOR OPTOPIA, IT IS USUALLY OVERLOOKED.
NEW PATH!
AAAAA
WITHOUT A11IES (4), THE DEMOGRAPHIC THAT WAS HOLDING ON TO THE ACCESSIBILITY ROPE IS NOW CUT OFF.
LET'S GET RID OF THE MAGPIE MENTALITY?
FOR YOUR NEXT FUN DATA SCIENCE PROJECT, INSTEAD OF SOME COMMUNITY-OVERFITTED IMAGE RECOGNITION CHALLENGE, MAYBE CHOOSE AN OPEN PROBLEM IN DIGITAL ACCESSIBILITY, SUCH AS AUTOMATIC VIDEO CAPTIONING. THEN HOPEFULLY ONE DAY THERE WILL BE "NO MORE CRAPTIONS" (5)
PART 2: GHOSTS IN THE SHELL
(WHO ARE WE BUILDING MODELS FOR?)
WE HAVEN'T YET FIGURED OUT HOW TO MAKE EXISTING DIGITAL PLATFORMS ACCESSIBLE TO EVERYONE, YET WE'RE ALREADY JUMPING TO FORGE A NEW "INTELLIGENT" CLASS OF WEB APPLICATIONS.
WE'RE SO CAUGHT UP IN THE "HOW" (USING ML/AI/DL/DS !!!) THAT WE FORGET TO ASK, "FOR WHOM"?
WHEN PLATFORMS ARE NOT DESIGNED FOR EVERYONE, THEY GIVE OFF THE STENCH OF "ENCODED INHOSPITALITY" (6).
SEEMINGLY INNOCUOUS THINGS SUCH AS POP-UPS AND EXPIRING FORMS ON WEBSITES COMPLETELY HIJACK THE ONLINE EXPERIENCE OF USERS WITH DISABILITIES WHO RELY ON SCREEN READERS.
Reality
90
Expectation

Feeding Information for Sustainability

The Green Consumption Assistant

Not having any sustainable options when buying a product online should no longer be an excuse for unsustainable consumption choices. The Green Consumption Assistant project sets out to support consumers in easily finding and buying sustainable products – by making use of the existing machine learning infrastructures in the retail industry.

The fundamental dilemma observed by the project team around Tilman Santarius and Maike Gossen from TU Berlin, Felix Biessmann from the Berliner Hochschule für Technik and the green search engine Ecosia was twofold. First, people say they want to make more sustainable choices but do not act on that desire when buying products. Second, the existing machine learning tools in the retail industry could be used to make sustainable consumption decisions a lot easier, but there is a lack of essential and comprehensive data about sustainable products to feed these systems.

Dimension: economic sustainability

Criteria: sustainability potential in application

Indicator: promotion of sustainable products

AI systems can be deployed in online shopping to promote more sustainable consumption through recommendation and search algorithms. Sustainable products can be given greater visibility in listings, product search results can highlight more sustainable alternatives and additional information can also be displayed, such as CO_2 emissions. AI systems should promote economization and sufficiency and encourage a shift away from unsustainable patterns of use (such as binge watching and food waste). Sustainability criteria like CO_2 emissions, working conditions and fairness must be programmed as relevant criteria in the decision-making process of the systems.

The solution: Building green databases and making sustainability aspects an essential criterium for an algorithm's automated decision-making. In the sphere of online shopping, automated recommender systems could then rank sustainable products more prominently than non-sustainable products. Transparency about such databases would likewise allow for systematic checks of what sustainability definitions and certifications are being used as the basis for a product's labeling as sustainable. That would ultimately empower consumers to make more informed choices.

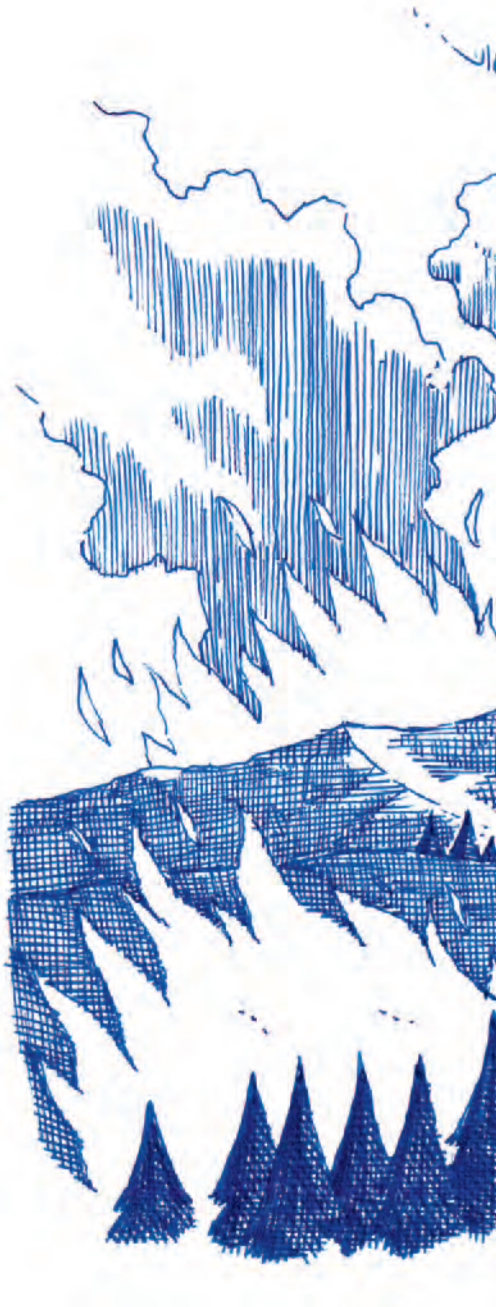
The Green Consumption Assistant addresses exactly this lack of green databases. The project team has been working on creating the GreenDB, a database containing sustainability information for consumer goods. The GreenDB is updated on a weekly basis and includes over 220,000 unique products from the largest online retailers in several European countries. In contrast to previous approaches to sustainability databases, the GreenDB covers only products that users are interested in: Its 26 product categories, currently mostly fashion and electronics, have been selected based on a careful analysis of the search logs of Ecosia users. The database displays information on the type of sustainability information that underlies any given product in the database – be it either more credible third-party verification or non-verified private sustainability labels. The database is used in the shopping tab of Ecosia's search site highlighting green products and thereby possibly encouraging consumers to make more sustainable choices. The GreenDB is publicly available and has two main purposes. One is for research. The other is for improving AI applications, such as recommendations and the reliability of sustainabilty information.

Beyond the usage of the GreenDB to drive sustainable consumption, it can also help to gain new insights into the availability of sustainable information on online fashion retail – and potentially infer appropriate policy changes. The relatively small ratio of only 14 percent of sustainability-tagged products in the online shops of Germany's largest fashion retailers are labeled with credible third-party verified sustainability labels. This underlines the difficulty faced by consumers in determining how sustainable a product is. The widely used private and non-certified labels prevent comparability and add confusion and uncertainty for consumers. More clarity and information are urgently needed, especially political initiatives tackling the risk of greenwashing resulting from uncertified and weak sustainability information.

Green Consumption Assistant

The *Green Consumption Assistant (GCA)* helps consumers in making online purchasing decisions that are more sustainable. It displays green product alternatives on the Ecosia search engine and provides information about more sustainable alternatives, such as references to repair, rental or sharing options. The basis for GCA's recommendations is a product database (GreenDB) of environmental and social sustainability information developed with the help of Machine Learning.

The *GCA* project is a partnership between the Technical University Berlin, the Berliner Hochschule für Technik and the green search engine Ecosia. This model project on the use of Artificial Intelligence in addressing environmental challenges has been funded by the German Environment Ministry.



Stuck in an Unsustainable Infrastructure

AI systems are not only data, nodes in a network or computational code – as popular visualizations would have us believe. They heavily rely on the exploitation of natural and social resources. The AI ethicist Aimee van Wynsberghe considers regulation to be the only way to hold Big Tech companies accountable for these hidden sustainability costs of AI.

In your paper on sustainable AI, you differentiate between AI for sustainability and the sustainability of AI. Why is it relevant to have a discussion on the sustainability of AI?

I wanted to remind people that there is a physical infrastructure required to make and use AI technology. And this physical infrastructure is currently unsustainable because it generates carbon emissions. The infrastructure required for the training and use of algorithms – things like batteries and microprocessors – require minerals, and the conditions under which humans have to work in in order to acquire these minerals are horrific. They are commonly referred to as blood minerals. There's also the issue of water and land usage to maintain this infrastructure. And what do we do with the electronic waste, the servers? We dump them in Asian countries and the people there have to suffer the environmental consequences. So, what I was trying to do with the distinction was to say it's not enough for us to say we will use this technology to achieve sustainability. We have to assess the technology itself for sustainability.

The discourse on AI for sustainability is progressing quite rapidly. Why isn't the discourse on the sustainability of AI keeping pace with it?

Once we uncover more of the hidden costs, we realize just how costly AI is, and then we have to burst the bubble all these companies are creating - Google, Amazon, Facebook, Apple. There's a strategic reason for them to hide these costs because Europe invested 3.2 billion euros in AI in 2020. The Big Tech companies try to avoid being required to measure the costs and to establish a complete understanding of their procurement chains. We

The Desirable Digitalization Project

The *Rethinking AI for Just and Sustainable Futures* research project examines AI development based on ethical principles. Led by Prof. Aimee van Wynsberghe, researchers involved in the project are working with the AI industry to develop sustainable and just principles for AI design and education. The project began in April 2022 and will run for five years.

Dimension: ecological sustainability

Criterion: indirect resource consumption

Indicator: key recycling figures

The production of computer hardware requires "conflict raw materials" or "conflict minerals," rare earths or precious metals the extraction of which is linked to human rights violations, appalling working conditions and environmental pollution. If the recyclables hardware contains are recovered at the time of disposal, they do not need to be extracted again for new hardware. Certified recycling specialists can separate hardware into the recyclable materials it contains and thus make it recyclable. Alternately, used hardware can be collected and reused by Original Equipment Manufacturers (OEMs) or refurbishment companies. They remove individual hardware components and reinsert them into used or new products. The percentage by weight of recycled or reused materials is an important metric for assessing the environmental sustainability of hardware.

don't see the discourse on the sustainability of AI progressing much because a) it would come with a lot more work for Big Tech, and b) we don't like to acknowledge that the hidden ecological costs are much more exorbitant than we think. The sheer complexity of the problem is an obstacle to the progression of the discourse. And furthermore, acknowledging the ecological costs of AI would curb our society's enthusiasm for the technology.

Do you expect the Big Tech companies to address the question anytime soon?

No. I think it really comes down to regulations. If we leave it up to the companies, we have to keep in mind that economics is their bottom line. They have an obligation to their shareholders, they have to make money. Having an AI ethics board was a strategic decision that made them look responsible, and this was basically their sole instigator. We need regulations that require Big Tech to evaluate the ecological impact of AI so that we can get the full picture of how bad the situation is.

How bad is the situation?

We have to act immediately. We are now in a state of "code red for humanity," as a recent IPCC report put it. Every technology we use needs to be evaluated for its impact on the environment. When it comes to AI, we've already reached a point that billions around the globe are invested in it, in every sector that you can imagine, and not just the algorithms but also the infrastructure being built on top of them. AI is spread out on a global scale, it's an incredibly pervasive technology. If we don't act now, it will be too late and we will be stuck in a kind of carbon lock-in caused by unsustainable infrastructure. The global community will bear the burden of the ecological costs.

We are often confronted with the claim that it's not easy for industry actors to be transparent regarding the sustainability of AI. How feasible do you think it would be for industry actors to design and produce sustainable AI?

I have frequently heard that imposing ethics on AI would stifle innovation by creating all these annoying checks and balances procedures. All we are doing is trying to push for good innovation that promotes social and ecological sustainability. There was a time when AI itself wasn't considered feasible, and now it's everywhere. So, why wouldn't implementing it in the right way be feasible? This is where regulation comes in. We need governments, the European Commission and the European Parliament to oblige Big Tech to measure and track carbon emissions and to look into the procurement conditions of minerals used in the infrastructure, regardless of the difficulty. If it's not possible to do it in a sustainable way, then don't do it at all. Otherwise, let's innovate. It's feasible if you push yourself to be innovative.

The AI Act currently being discussed at the European level is supposed to minimize risks stemming from AI and protect fundamental rights. What's your stance on the AI Act in its current form?

In the sustainability department, it's not doing anything at all really. My biggest problem with the AI Act is that it doesn't conceptualize ecological risks as risks that require a risk assessment. We need more transparency on the sustainability of AI. This would be the first step towards a discussion about a carbon cap or a training cap allowing for a certain number of GPUs in algorithm training for a certain number of hours. Before we have the data on how much electricity is needed to train an algorithm or how much water is needed to cool down the servers, formulating demands would be making uneducated guesses. That's why I advocate for mandatory measuring and transparency.

Dimension: economic sustainability

Criteria: working conditions and jobs

Indicator: fair wages along the value chain

Problematic working conditions in the development of AI exist not only in hardware production, but also in the preparation of data. The data sets needed for training AI systems usually must first be *labeled*, i.e., classified and annotated by crowd- or clickworkers. These workers often perform small tasks (per click) under precarious conditions for companies without being formally employed. When developing as well as purchasing AI, care should be taken to ensure that working conditions are fair throughout an AI's entire lifecycle. This includes adequate pay, good working conditions and opportunities for advancement along with further training, even for clickworkers.

PROF. DR. AIMEE VAN WYNSBERGHE

… is the Alexander von Humboldt Professor for Applied Ethics of Artificial Intelligence at the University of Bonn in Germany. She is Co-Founder and Co-Director of the *Foundation for Responsible Robotics* and serves as a member of the European Commission's *High-Level Expert Group on AI* as well as a member of the World Economic Forum's *Global Futures Council on Artificial Intelligence and Humanity*. She is the recipient of a Dutch Research Council personal grant for the study of the responsible design of service robots. Van Wynsberghe is a 2018 *L'Oreal UNESCO For Women in Science* laureate, a founding Board Member of the *Netherlands AI Alliance*, a Founding Editor of the international peer-reviewed journal *AI & Ethics* (Springer Nature), and author of the book "Healthcare Robots: Ethics, Design and Implementation."

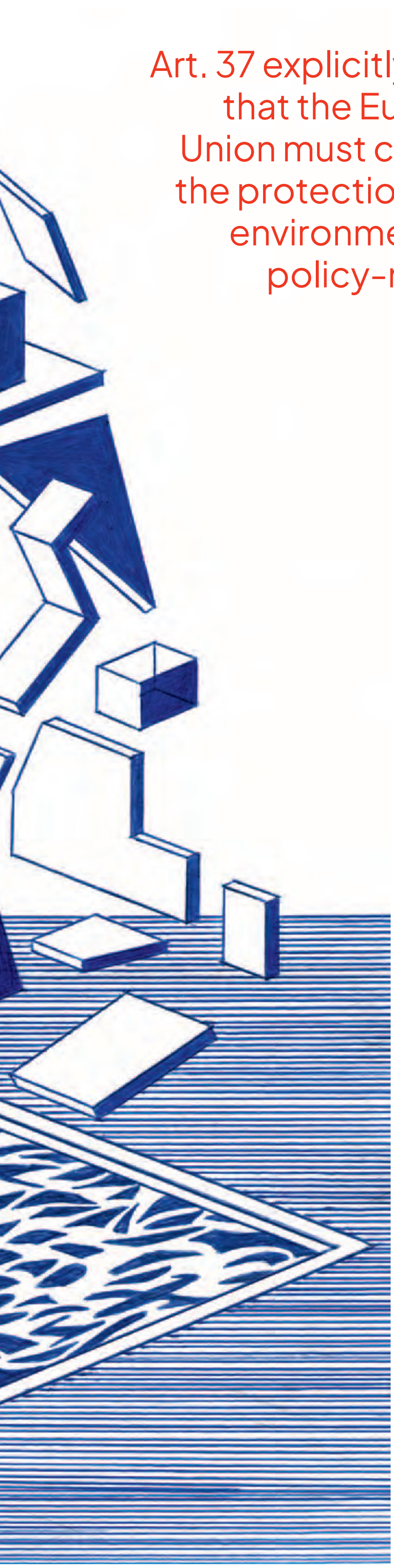
The EU's AI Act:

Dangerously Neglecting Environmental Risks

The EU has recognized the need to address risks associated with AI on a political level. But when it comes to the technology's resource consumption and environmental impacts, the AI Act is turning a blind eye.

In April 2021, the European Commission published its proposal for the Artificial Intelligence Act (AI Act) – as a response to the increasing need to regulate the technology. Even though the Act claims to have the consideration of fundamental risks of AI on societies and individuals at its heart, the proposal is a disappointment regarding the environmental risks of AI. This is all the more upsetting since the Commission's White Paper on Artificial Intelligence, which preceded the Act, explicitly pointed out that AI development must proceed in an environmentally friendly way.

The AI Act proposes a regulation laying down harmonized rules for the protection of safety, health and fundamental rights against potential harms stemming from AI, while at the same time fostering innovation. In order to achieve its goal, the AI Act takes a risk-based approach, setting rules based on the perceived level of risk of AI systems or of their deployment. But the AI Act fails to account for the environmental risks stemming from the development and deployment of AI systems.



Art. 37 explicitly states that the European Union must consider the protection of the environment in its policy-making.

In our view, if the AI Act's aim is to protect our safety, health and fundamental rights, it would be negligent for the European Union to not account for the protection of the environment. The European institutions can hardly be in doubt about the environmental risks of AI systems, be it the tremendous resource consumption associated with some AI systems or their underlying infrastructures, when there is a plethora of evidence available. In its initial proposal, the Commission thus did not do justice to the Charter of Fundamental Rights of the European Union, which in Art. 37 explicitly states that the European Union must consider the protection of the environment in its policy-making.

Following the promising start in the White Paper on Artificial Intelligence, the draft AI Act is a disappointment. The White Paper stressed that the "environmental impact of AI systems needs to be duly considered throughout their lifecycle and across the entire supply chain, e. g. as regards resource usage for the training of algorithms and the storage data." But in the current draft of the AI Act, there are no environmental mandates placed on providers and/or deployers. As it currently stands in the AI Act, providers *may* create and apply codes of conduct, which can include voluntary commitments regarding environmental sustainability. But voluntary applications of codes of conduct can hardly be considered an adequate response to an increasingly pervasive and resource-intensive technology such as AI.

Thus, as currently written, the AI Act so far misses a crucial opportunity to ensure that the development and use of AI systems is done in a sustainable, resource-friendly manner in which our planetary boundaries are respected. This shortcoming is at odds with our collective endeavor to combat climate change as well as with the objectives of the European Green Deal and other policy initiatives of the EU. The European institutions are still negotiating the Act, so there is time to correct this neglect of one of the most central risks of AI technologies. A necessary first step would be for the AI Act to acknowledge the vast environmental risks of AI by making them a relevant criterion for assessing whether AI systems should be classified as high-risk or not. Consequently, organizations developing and implementing AI systems should monitor their AI-related resource consumption, be required to make such data transparent and take adequate steps to develop and deploy AI in an environmentally friendly way.

We need more insight into the actual resource consumption of AI in order to establish a more evidence-based regulation of AI technologies. The AI Act provides an opportunity, which should not be missed. It is now up to the European Parliament and member states to compensate for the Commission's ommission.

NIKOLETT ASZÓDI

… is a Policy and Advocacy Manager at AlgorithmWatch. Her work focuses on the use of automated decision-making (ADM) systems in the public sector and on horizontal EU regulations in the field of ADM – in particular the EU AI Act.

Ecologically Sustainable AI Requires Regulation

Artificial Intelligence is seen as a key technology of the 21st century. It can serve countless practical purposes: translations, medical diagnoses, personalized product recommendations and much more. As such, it is likely that AI will gradually enter nearly all areas of society, not only in the form of new products and services, but also through the improvement of existing processes, making them "smarter." The increasing saturation of society with AI-based solutions, however, means that global energy consumption will increase, not only because of end device usage, like smartphones. A significant portion of the energy consumption associated with AI applications takes place externally, through data transfers and at data centers. It is true that data-center energy efficiency is increasing steadily. But if the sector remains largely unregulated, a rebound effect could set in, meaning the cost reductions associated with energy savings could lead to more intense usage, and therefore to a growth in absolute energy consumption. The European Commission estimates that data-center energy consumption in the EU will increase from 77 terawatt hours (2.7 percent of overall electricity consumption) in 2018 to 99 terawatt hours (3.2 percent of total energy consumption) in 2030. In short: The growing number of data transfers and rising amount of data processing translate into more energy consumption.

It is thus important to consider how AI solutions can be developed, controlled and used as energy efficiently as possible. One potential approach is to educate consumers about AI's ecological footprint. Certificates and labels like those the European Union provides for "Green IT" could help consumers recognize and choose environmentally friendly options. It's also feasible to require so-called "carbon impact assessments" for the development and sale of AI solutions and hardware. If the ecological footprint of AI applications is quantified and presented publicly in this way, it could influence consumer behavior.

But it is unclear how effective information is on its own. Not only does this create yet another area where consumers must learn to make informed choices, but experiences in other sectors have also shown that the combination of sufficient knowledge and an environmentally friendly disposition does not necessarily lead to a change in behavior. And energy-efficient applications are unlikely to catch on if they are more expensive or have lower performance, as shown by the results of the study we conducted called "*Consumers are willing to pay a price for explainable, but not for green AI. Evidence from a choice-based conjoint analysis*." Especially when

DR. MARKUS B. SIEWERT

... is executive director of the TUM Think Tank at the Munich School of Public Policy. As a political scientist, he is interested in socio-political challenges in relation to areas of digital transformation, Artificial Intelligence and sustainability, as well as the socio-political governance of these areas.

it comes to free systems, people are unlikely to accept fees for greater energy efficiency standards. Environmentally friendly AI is therefore unlikely to establish itself on the market based on information and labels alone. That means that political action and government regulation will need to be implemented at points prior to consumer decision-making, as the following graphic shows.

Financial instruments can focus on the marketing stage and regulate supply and demand through pricing. A relatively simple and easily implementable solution would be the inclusion of both producers of AI applications and products as well as data centers in the existing CO_2 emissions trading system. This would result in stronger incentives for the development and usage of energy efficient applications and infrastructures for information. By making externally consumed energy and the resulting emissions more expensive, companies would be driven to consider environmental sustainability at the development stage ("Green AI") instead of only looking at performance ("Red AI"). Such measures can make an important difference. An OECD report shows that different technical decisions - regarding the choice of model, hardware and data center, including their locations - can lead to enormous energy savings.

Finally, regulation can also focus on the development stage of AI applications. This could happen, for example, through the top-runner approach as used in Japan: All providers need to reach the highest energy efficiency standards of the leading provider within a predetermined timespan. Otherwise, they risk being prevented from offering their products on the market – through a government ban, for example. This kind of hard regulation sets new standards by dynamically linking market competition to the respective state of the technology.

Ideally, the use of all of these instruments would complement each other and would be accompanied by the continuous collection and evaluation of data. In that effort, state regulators should not be forced to rely exclusively on reports and information supplied by the companies themselves. The government must also develop the capacity to effectively monitor companies and, should it become necessary, to intervene.

DR. PASCAL KÖNIG

... is a Political Scientist at the Technical University of Kaiserslautern and a Visiting Scholar at the *Minda de Gunzburg Center for European Studies* at Harvard University. His scholarly interests include the consequences of the datafication of societies, governance by algorithms (including ethical and regulatory aspects) and how political actors deal with the challenges of digital transformation.

PROF. DR. STEFAN WURSTER

... is a Professor of Policy Analysis at the Technical University of Munich (TUM). His research focuses on policy comparisons in political fields with a strong link to sustainability, such as education, research, innovation, environmental policy and energy policy. His primary research interests are different instruments of political governance and comparisons between democracy and autocracy.

Modular AI

The software company elevait develops AI products for the digitalization and automation of business processes. The system's modular design reduces training time and saves resources. In addition, it enables small and medium-sized companies to make use of AI solutions.

How did elevait decide to focus on the development of sustainable AI?

In the years since our founding, we have seen the ways in which AI implementation doesn't work well: Development should not be oriented toward the projects of individual customers with specific problems. One-off solutions to such problems cannot be transferred to other projects, meaning that the time and money was not exactly invested sustainably. Instead, we now provide companies that want to get specific results from their data with the appropriate AI building blocks, enabling them to process their customer data without having to start from scratch each time. There are, of course, limits to this principle, but it has proven to be very practical and sustainable – not only economically, but also environmentally. If a new model must be trained for each AI project, the server requirements and energy needed for computation are enormous. The less training we require, the fewer resources are consumed to build the models and thus deliver solutions to the customer.

Can you explain this building block principle? How can AI be constructed such that it is adaptable and useable for a variety of small and medium-sized enterprises?

Imagine a project where documents need to be extracted. If we were to develop an AI solution specifically for this project, we would have to, for example, build a model that can recognize the document type, extract certain areas and read the corresponding text. In our approach, the tasks in this process are treated as modular building blocks. There is a defined model for each step of the work – for handwriting recognition, for example. This module is not, or only slightly, dependent on the specific project profile of each customer. We can use these types of trained models in all use cases where they are needed. We have trained models for specific small tasks according to this building-block principle, and we then arrange the individual modules using a workflow, which is based on the project requirements.

With this modular process, companies are no longer involved in the development of AI themselves. Is it not problematic if the companies don't even know what is happening and how?

We have recognized that we can't offer AI software to individual customers with the expectation that we will sell it and be done with it. The implementation of AI follows certain workflows that are established within the company. These workflows evolve over time – also in the sense that the model quality improves with the trained elements. Trust must be established with clients. We work with transparent benchmarks and tell them, for example: We generated a representative data set and analyzed it with the models; 80 percent of the documents were recognized, and the data was read with 95-percent accuracy. Customers can then give us feedback as to whether these values are satisfactory for them. In the

end, if the results are right, it doesn't matter if the AI solution is a black box. But we must first get to the stage where they trust this black box.

What types of use cases are companies approaching you with?

Our core business is the development of AI products for the automation of business processes, like automated document processing. For example, automated capture and processing of forms for ordering or inventory, medical history forms in hospitals, invoices, delivery bills, service tickets, construction plans or pure text in the broadest sense.

What is the impact of automation on companies?

One of our customers is a manufacturer of orthopedic products, to take one example. Individual orders make up a significant portion of its business. Patients have their legs or arms measured individually in a health care supply store and these measurements are then transmitted to the supplier, which must manufacture and ship the required product within a short period of time. Such a process has an extremely tight timeline and a high degree of individualization. Everything has to fit perfectly as well. Up to 40 measurements along with 40 to 80 configuration options must be considered for each order. To date, up to 80 percent of the orders are processed through a form sent in by email or fax. Our customer receives several thousand order forms each day, often filled out by hand. It used to be the case that around 40 employees would type in the forms in two shifts each day. But it was growing increasingly difficult for our client to find employees for this task, and that contravened the company's growth plans. They wanted to process more orders each day but doing so wasn't possible with the processes in place. Our task was to recognize the orders in a partially or fully automated way and to enter them into the system.

Dimension: environmental sustainability

Criterion: energy consumption

Indicator: consideration and optimization of energy efficiency

Considerable energy can be saved in the development and deployment of AI through, for example, compressing models, efficient model training, the adoption of data-minimalist approaches, using pre-trained models, using less-complex models or using efficient software and hardware infrastructure. There is still limited awareness about this issue and a lack of expertise on the appropriate methods. Still, organizations that develop or deploy AI should make the energy efficiency of an AI system a central criterion in their decision-making and selection processes to act in an environmentally sustainable manner.

When automation processes are introduced, there is always concern that jobs will be lost as a result. How did you handle those fears?

We have to bear in mind that our society is facing a demographic change in which a great many people will be leaving the workforce in the next 10 years. At the same time, the number of documents, information and data that need to be processed will steadily increase.

In the health care supply store, patients' legs and arms are being measured and written down by hand. This data is then sent - sometimes even by fax - to the production site.

AI is one solution for addressing this ever-widening gap between available workforce and the need for processing data. This isn't about taking away people's jobs, but about addressing this problem. In addition, AI should make work easier, and employees should be deployed where AI cannot provide assistance.

Still, the use of AI is changing the work of employees. If it makes data entry faster and more efficient, what will become of the employees who were hired to do that job?

When you attempt to automate a process, it's not enough to just use AI. A smooth process flow hinges on the specialized knowledge and experience of employees. For example, employees have known customer X for a long time. For years, when this customer orders something, they have always added an extra 10 percent to the specified dimensions. This is because they know from experience that this customer's orders are always too tight, and without adjustment, he will change them later, anyway. For an automated process to be carried out successfully, such nuances must also be included in the target system. Mere data extraction isn't enough. As such, along with the AI system, we also provide our customers with an editor they can use to define rules themselves. It allows you to specify in the system: If a specific customer number and the following attribute is recognized, then please delete that value or add 10 percent to all measurements, for example. We have built the tool in a way that our customers can make such edits themselves and consider what the system is still missing. In this way, they can gradually incorporate the existing specialist knowledge of individual employees into the system, bit by bit, to steadily increase the degree of automation.

To what extent do workers need further training for that purpose?

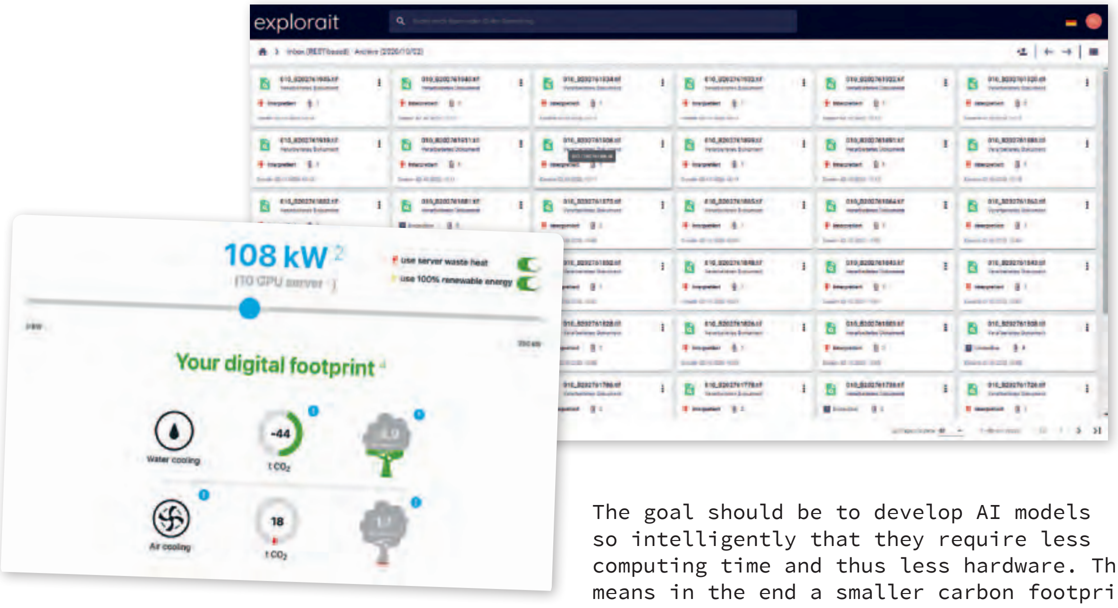
A workshop was held with employees from the customer service department – those people who had previously typed in the data by hand – to explain how the rules could be adapted. They now maintain the system under the

Dimension: economic sustainability

Criteria: working conditions and jobs

Indicator: evaluation and optimization of working conditions

AI systems are increasingly used in the workplace, which can either improve or worsen working conditions. Many fear that the use of AI will lead to job losses. Organizations should assess the consequences for workers before implementing an AI system. The systems can, for example, lead to work becoming more monotonous, to employees being more closely monitored or to workers' qualifications for certain activities losing their value. Interests must be balanced, through offers of additional training, for example.



The goal should be to develop AI models so intelligently that they require less computing time and thus less hardware. This means in the end a smaller carbon footprint.

supervision of the head of customer service and an IT manager. But the rules come from the employees. The focus of their work is also shifting as a result of the use of AI. They now spend their time checking and processing borderline cases that can't be handled by the system and have to be dealt with by the employees – who, in turn, now have more time to devote individually to these cases.

You work together with the sustainable data center Cloud&Heat. How important is it to consider resource conservation already in the development phase of AI?

There is a strong interaction between economic and environmental factors. For us, a key question is how to achieve the lowest possible computing time. Can I reduce the computing time simply by the choice of the architecture of my AI model and the software behind it? This is a very simple and important factor in being environmentally sustainable, because it means we use less electricity. And it has the economic advantage that it costs less. We also can't forget about the hardware that is necessary for the calculations. The construction of the hardware already generates a large carbon footprint. We need to train our models on GPUs. If I need 100 GPUs, the footprint is correspondingly large. But if I choose the architecture of my model intelligently enough to compute a similar result on 10 GPUs, then I generate a much smaller carbon footprint. We are intrinsically motivated to optimize our models, but at the same time, there are a number of monetary incentives as well.

GREGOR BLICHMANN

... conducted research on web-based software and service developments as a Research Assistant at the TU Dresden after his studies in computer science. The software engineer and software architect then joined the predecessor company of *elevait*, where he helped construct its basic software architecture. With the launch of *elevait* and the growth of the company's technical teams, he increasingly became responsible for coordination between the various teams until he ultimately rose to become the company's Chief Technology Officer (CTO).

Putting Excess Heat to Use: How to Turn Data Centers into Sustainable Radiators

Dresden-based startup *Cloud&Heat* is looking for ways to make cloud infrastructure more sustainable. With its data center cooling system, which relies on direct hot water cooling, the company can save up to 710 tons of carbon dioxide annually relative to traditional air-cooled centers, according to a model calculation performed as part of a pilot project in Frankfurt. Ronny Reinhardt, Team Lead of Business Development at *Cloud&Heat*, explains how waste heat from data centers can be used for heating and thus make the cloud more sustainable.

Hardware comparison: The more energy-efficient water cooling system (on the right) not only requires less energy, but also makes the waste heat generated usable – for example to heat buildings.

What role do data centers play in AI resource consumption?

AI consumes computing power and thus energy. The AI models usually run on GPU hardware, graphics cards that have a very high power consumption. When we compute on servers, in terms of energy, it is really only electrical energy that is converted into heat. As a consequence, this means that additional energy is required to cool the servers. In German financial center Frankfurt alone, for example, the heat generated in data centers could theoretically heat the entire city. Data centers as the basis for cloud and AI solutions constitute a rapidly growing industry that we need to be thinking about now to ensure we are well positioned for the future.

© Cloud&Heat

Server racks: Cloud&Heat is trying to curb the cloud and data center industry's rising resource consumption and reduce its carbon footprint.

Is it sufficient for data centers to rely on green power or renewable energy and, if necessary, offset their CO_2 emissions?

From our perspective, it's not enough. Renewable energy, of course, is better than a conventional electricity mix. But our fundamental goal should be that of using as little energy as possible. This starts with software development at the applications level and extends through the various software levels right up to the data centers. The question, then, is this: How can a data center be operated as efficiently as possible? Cooling is a major factor. Ten years ago, the same amount of energy necessary to run the servers was needed to then cool them. Today, we only need around 20 to 30 percent. Still, we must continually improve. The second major issue is waste heat utilization. As I mentioned, from an energy point of view, we are basically inputting electrical energy into data centers and getting heat out – even if, of course, there is some computing that takes place as well. This waste heat needs to be put to good use, and there are many different approaches for doing so, such as feeding it into the heat supply of buildings or connecting it to district heating networks.

What measures do you use to optimize energy efficiency in the data centers?

Our focus is on what is called direct hot water cooling. We siphon off the heat produced by the components of the server – the processor or the graphics cards – directly to utilize it for other purposes, such as heating a building, for example, as we are doing in our pilot project in the former data center of the European Central Bank in a high-rise building in Frankfurt. This is a technical challenge, because we must tap the heat at a high temperature – otherwise, little can be done with it. Direct hot water cooling is more energy efficient than air cooling, which requires air to be cooled using traditional cooling systems. Doing so consumes significantly more energy than simply running a pump that circulates water, allowing it to flow over the servers and the hot components. That already results in lower energy use. Second, we then use the waste heat by feeding it into the building – into the heating system, for example. On the basis of this technology, we are deploying an open source-based cloud solution that AI customers can use to perform AI model training and inference in the most sustainable way possible.

Together with Vattenfall, you have also launched a pilot project in Sweden to use waste heat for energy supply. How does that work, exactly? And what is the goal of the project?

Together with Vattenfall, we are building a cloud for AI companies or for other firms that require high computing power and for which the issues of sustainability and energy efficiency are vital. We set up our data center containers with direct hot water cooling on the site of a biomass power plant. We are, in other words, right at the renewable energy source. Our facilities there are so efficient that we only need about seven percent additional energy to run the data center relative to the immediate server power. We are also connected to the district heating network, into which we feed the heat generated during the cooling process. It is then routed onward to surrounding households near Stockholm.

You mentioned at the beginning that the heating needs of the city of Frankfurt could theoretically be met by the heat generated in data centers. Why isn't that being done? What are the specific hurdles?

Classic data centers are still operated with air cooling, and this makes waste heat utilization difficult, because the temperature level of the air is not particularly high and heat transport over long distances is difficult. To change the status quo with air cooling, a great many stakeholders would need to coordinate and move forward together. The large data centers are mostly colocation data centers where customers rent access to the servers. So, customers would have to bring in water-cooled hardware themselves, and the data centers would have to create the necessary infrastructure. This is the reason that *Cloud&Heat* exists in its current form. We have our own data center, and we can offer customers our own cloud. The biggest obstacle is that a great many stakeholders in the market tend to think conservatively in this area. To retool, data centers must first invest and recoup the investment costs through savings in operating costs or compensation provided for waste heat.

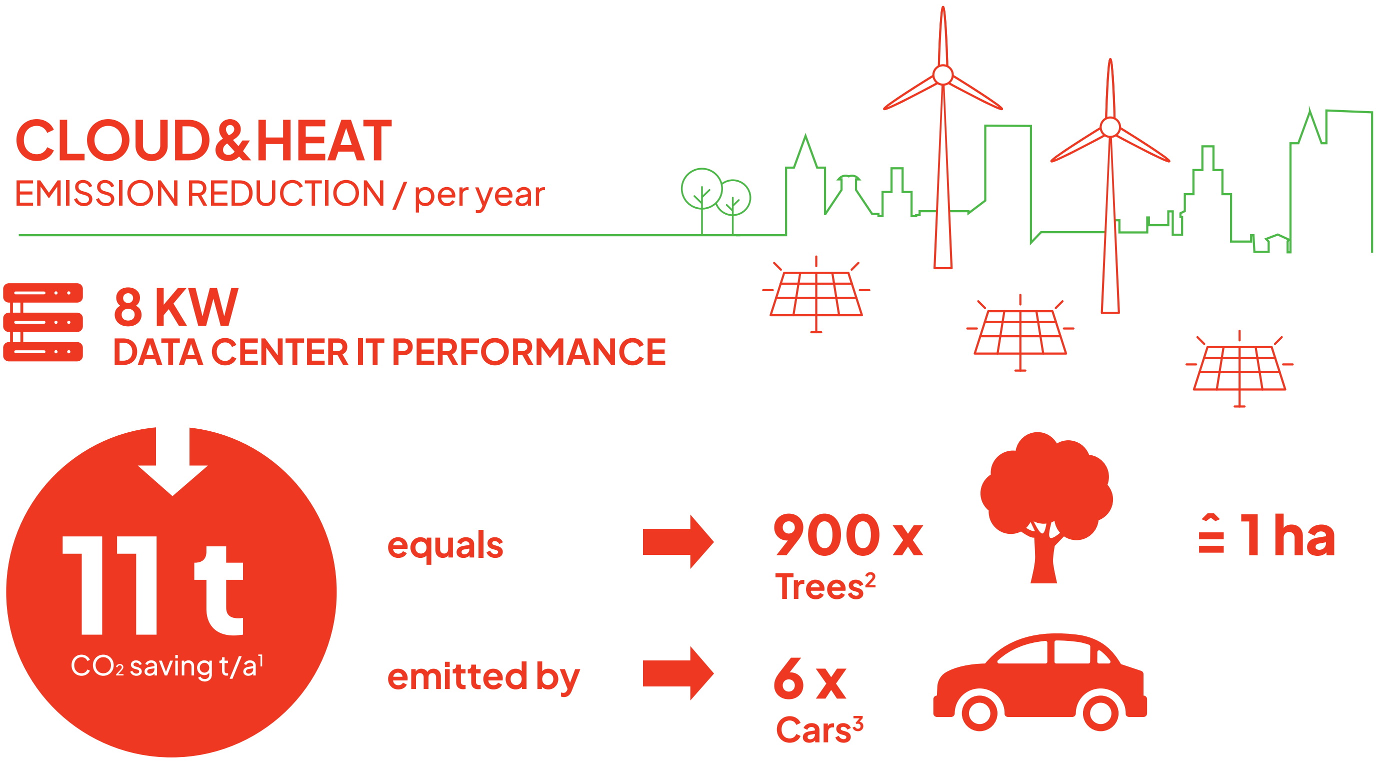
Are there other barriers preventing a greater number of data centers from adopting water cooling or other innovative and energy-efficient cooling systems?

First, it is also a question of habits and availability. There is still far more air-cooled hardware out there than water-cooled hardware. Although the number of providers and models is increasing, the breadth of offerings hasn't yet reached the same level. Some say it is the manufacturers who need to make a move. The manufacturers, though, say that demand for water-cooled systems isn't great enough for them to switch their product lineups. So progress is generally only being made in small increments. That's why we are providing support on this front and have, for example, developed a water-cooled server together with Thomas-Krenn. AG, which was one of the first systems to receive the Blue Angel label for environmentally friendly products.

There are growing calls for the provision of waste heat to be subsidized.

To promote the shift to more sustainable data center infrastructure, European countries are testing a variety of subsidy models. The Netherlands Enterprise Agency (Rijksdienst voor Ondernemend Nederland), for example, subsidizes the introduction of technologies that can reduce greenhouse gas emissions at the behest of the Dutch Ministry of Economic Affairs and Climate. If companies or organizations provide waste heat, they can receive a subsidy of €0.033 to €0.044 per kilowatt hour. Such subsidies can provide important incentives for data centers to make greater use of waste heat.

Source: https://english.rvo.nl/sites/default/files/2020/11/Brochure%20SDE%20plus%20plus%202020.pdf



*Source: 1) Cloud&Heat cooling technology vs. traditional air-cooled data center. 2) https://www.co2online.de/service/klima-orakel/beitrag/wie-viele-baeume-braucht-es-um-eine-tonne-co2-zu-binden-10658/ 3) VW Passat BlueMotion or BMW 320d; standard consumption of 1.92 CO_2 t/car at 15,000 km annual mileage.

A data center with a total IT power of 8 kW can save around 11 tons of CO_2 per year by using water cooling instead of air cooling. This is equivalent to the emissions emitted by six cars per year. Compensating for these emissions would require one hectare of forest (with 10,000 m2) or 900 trees. Source: https://www.cloudandheat.com/wp-content/uploads/2020/02/2020_CloudHeat-Whitepaper-Cost-saving-Potential.pdf

It would make perfect sense for society to create an incentive to reward the extraction of CO_2-free heat with a premium. Then the necessary technologies would be developed more quickly.

The third point is transparency toward users. On the one hand, this concerns the use of the cloud by companies. The fact that certain computing steps cause a certain amount of CO_2 emissions needs to be transparently communicated. But end users also lack any awareness of the fact that using certain apps or services produces CO_2 emissions. Awareness of such CO_2 footprints has already developed in other areas.

There should also be greater political support for an issue that is so important to society. There are models from other countries in which, for example, the waste heat discharged from such processes is remunerated. Essentially, it is CO_2-free heat, since the CO_2 has already been “consumed” during the computation step anyway. So, we can either destroy this heat or put it to good use. It would make perfect sense for society to create an incentive to reward the extraction of CO_2-free heat with a premium. Then the necessary technologies would be developed more quickly.

Companies often have no data at all on the CO_2 footprint and energy consumption of AI systems. It is claimed that the amount of work necessary for the collection of such data is simply too great. How difficult is it to create transparency in this area?

It is, in fact, not easy. In the cloud context, for example, different customers share a server, so it isn't immediately possible to say how much customer X consumed on the server or how much energy was consumed in this or that computing process. But there are solutions for that, too. We are addressing precisely these kinds of questions in the context of a major European research

project (IPCEI-CIS). The important thing here is to arrive at the solution as quickly as possible, even if a high degree of accuracy cannot be immediately achieved. Even the CO_2 footprints that are found on food products aren't accurate down to the decimal place. That's not really the point, though. What is important is that we have an informed sense that some foods lead to large CO_2 emissions and others do not. By the same token, we also need to take a step forward with cloud solutions to create more transparency and, looking ahead, move computing tasks to the point where CO_2 is lowest.

Does focusing on environmentally sustainable solutions also make economic sense?

I believe we will see increasing convergence here. We are seeing right now that as electricity prices rise, it is becoming increasingly attractive to improve energy efficiency. In some countries, the cost of electricity is still so low that no one has to worry much about it. In such places, the environment and the economy are diverging. Of course, the solution isn't to raise the price of electricity as an innovation incentive for energy-efficient technologies. That would just result in data centers relocating to other countries. At the same time, the environment and the economy are no longer mutually exclusive.

In its coalition agreement, the German government states that new data centers will have to be operated in a climate-neutral manner from 2027. What do you think of this announcement?

It is good that the problem has been recognized at the political level, but it is also extremely important to act quickly, because megawatt-scale data centers are now being planned and built. If we don't take countermeasures now, the data centers will operate as designed for decades. The local government of Amsterdam no longer allows new data centers to be built unless they include a plan for waste heat. So, there is already movement. But governments continue to struggle with setting criteria aimed at sustainability.

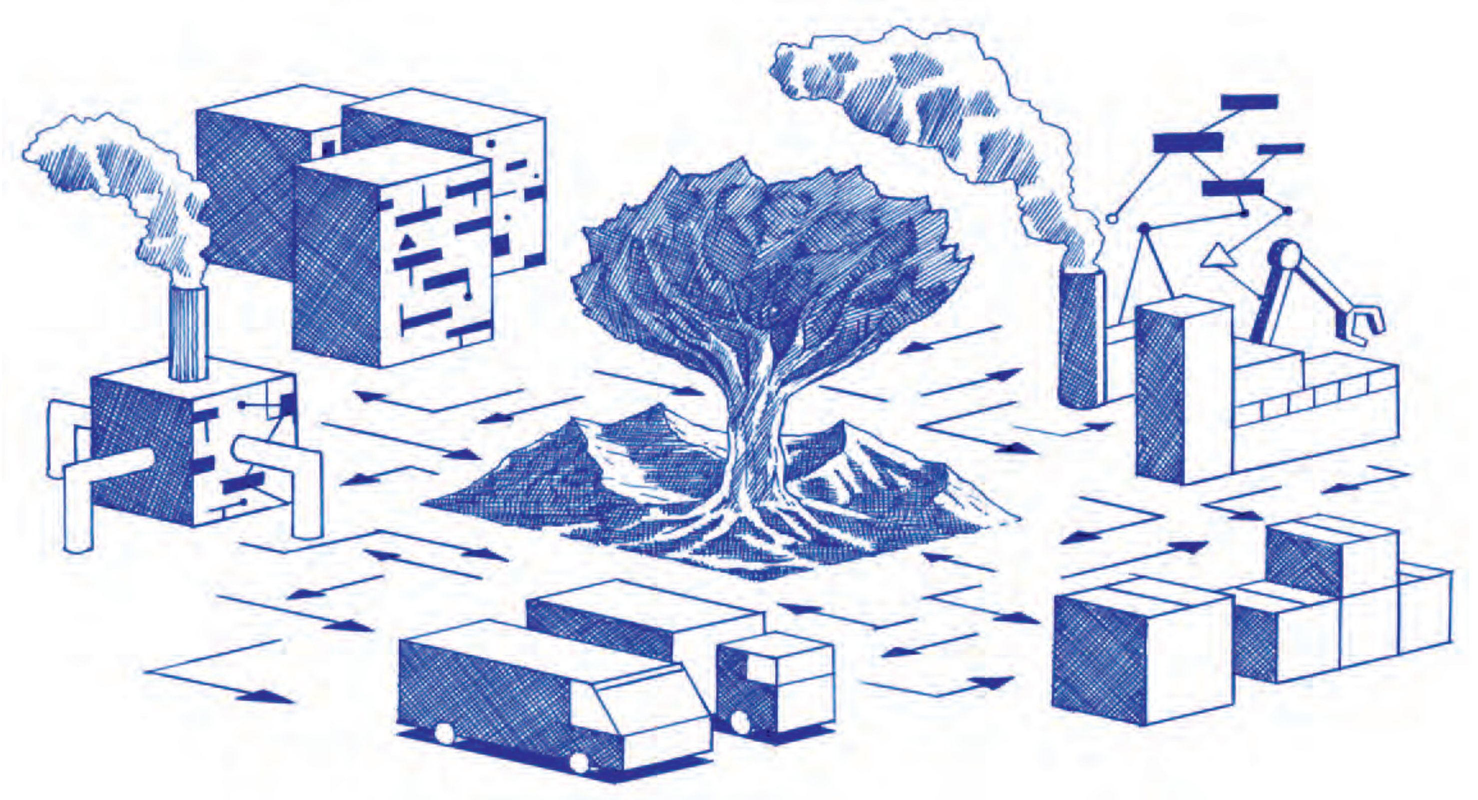
What do you think of environmental labels for data centers?

This is a good approach for establishing a certain framework that everyone can adhere to. But I don't think there are many data centers out there yet that can meet the requirements. However, we do orient ourselves on these kinds of ambitious goals. In addition, we also need to make sure that there is enough breathing room to allow new solutions that are not yet reflected in any sustainability label to develop their full impact.

If we don't take countermeasures now, the data centers will operate as designed for decades. The local government of Amsterdam no longer allows new data centers to be built unless they include a plan for waste heat.

DR. RONNY REINHARDT

… serves as Team Lead Business Development at *Cloud&Heat Technologies*. He is actively engaged in the European cloud and data initiatives Gaia-X and IPCEI-CIS. For Gaia-X, he served as a member of the Technical Committee, and he is now one of the coordinators of the GREEN-CIS consortium at IPCEI-CIS. Reinhardt is also a member of the Climate Change Working Group at the German AI Association, and he is involved in the Large European AI Models (LEAM) initiative. Previously, he conducted research and taught technology and innovation management at FSU Jena, the University of Utah and TU Dresden.



Coming in Our Next Issue of

sustain

In our next issue of the magazine, we will take a deeper look at the world of coding, delving into pressing topics including:

▶ How coding can boost resource efficiency

▶ Methods to foster fairness in AI systems

▶ How AI creates a more sustainable energy supply in buildings

We will provide guidelines for sustainable AI and step-by-step instructions for sustainable solutions throughout the entire life cycle of an AI system – from the initial idea to the application.

Imprint

Sustain Magazine #1
June 2022
Available online at https://algorithmwatch.org/en/sustain-magazine-2022

Publishers
AW AlgorithmWatch gGmbH
Responsibility (according to § 5 TMG and § 55 RStV): Matthias Spielkamp
Linienstr. 13
10178 Berlin
Germany

Content Lead
Anne Mollen

Editorial Lead
Waldemar Kesler

Content Support
Andreas Meyer (DAI Laboratory, TU Berlin)
Friederike Rohde (Institut für ökologische Wirtschaftsforschung)
Josephin Wagner (Institut für ökologische Wirtschaftsforschung)

Illustrations
Juliette Baily, Kevin Lucbert

Media
Marc Thümmler

Design / Layout
Beate Autering (AD), Benito Felkel, Tiger Stangl

Translation and Copy Editing
Dacha Media GbR: Daryl Lindsey, Charles Hawley
Additional translations by Breandáin O'Shea

Output from the project
"SustAIn: The Sustainability Index for Artificial Intelligence"

sustain
Sustainability Index
for Artificial Intelligence

Project partners:

ALGORITHM WATCH

i ö w
INSTITUT FÜR ÖKOLOGISCHE WIRTSCHAFTSFORSCHUNG

DAI-Labor
TU Berlin

Supported by:

Federal Ministry
for the Environment, Nature Conservation,
Nuclear Safety and Consumer Protection

based on a decision of
the German Bundestag

sustain

Sustainability Index
for Artificial Intelligence

Project Partners:

AlgorithmWatch

AlgorithmWatch is a non-profit research and advocacy organization whose aim is to monitor and analyze automated decision-making (ADM) systems and their impact on society. One pillar of our work is addressing ADM systems and sustainability. https://algorithmwatch.org/en/

ALGORITHM
WATCH

Institute for Ecological Economy Research (IÖW)

IÖW is a leading scientific institute in the field of practice-oriented sustainability research. We devise strategies and approaches for viable, long-term economic activity – for an economy which enables a good life and preserves natural resources. We have been dealing with issues of the future for more than 30 years, consistently finding new and frequently unusual answers. https://www.ioew.de/en/

i | ö | w
INSTITUT FÜR ÖKOLOGISCHE
WIRTSCHAFTSFORSCHUNG

Distributed Artificial Intelligence Laboratory

The DAI-Labor at the Technical University of Berlin views itself as a mediator between university-driven research and industrial applications. With our interdisciplinary team, we create innovations and translate university research into applications for everyday life in close cooperation with other scientific and industrial institutions.
https://dai-labor.de/en/

DAI-Labor
TU Berlin