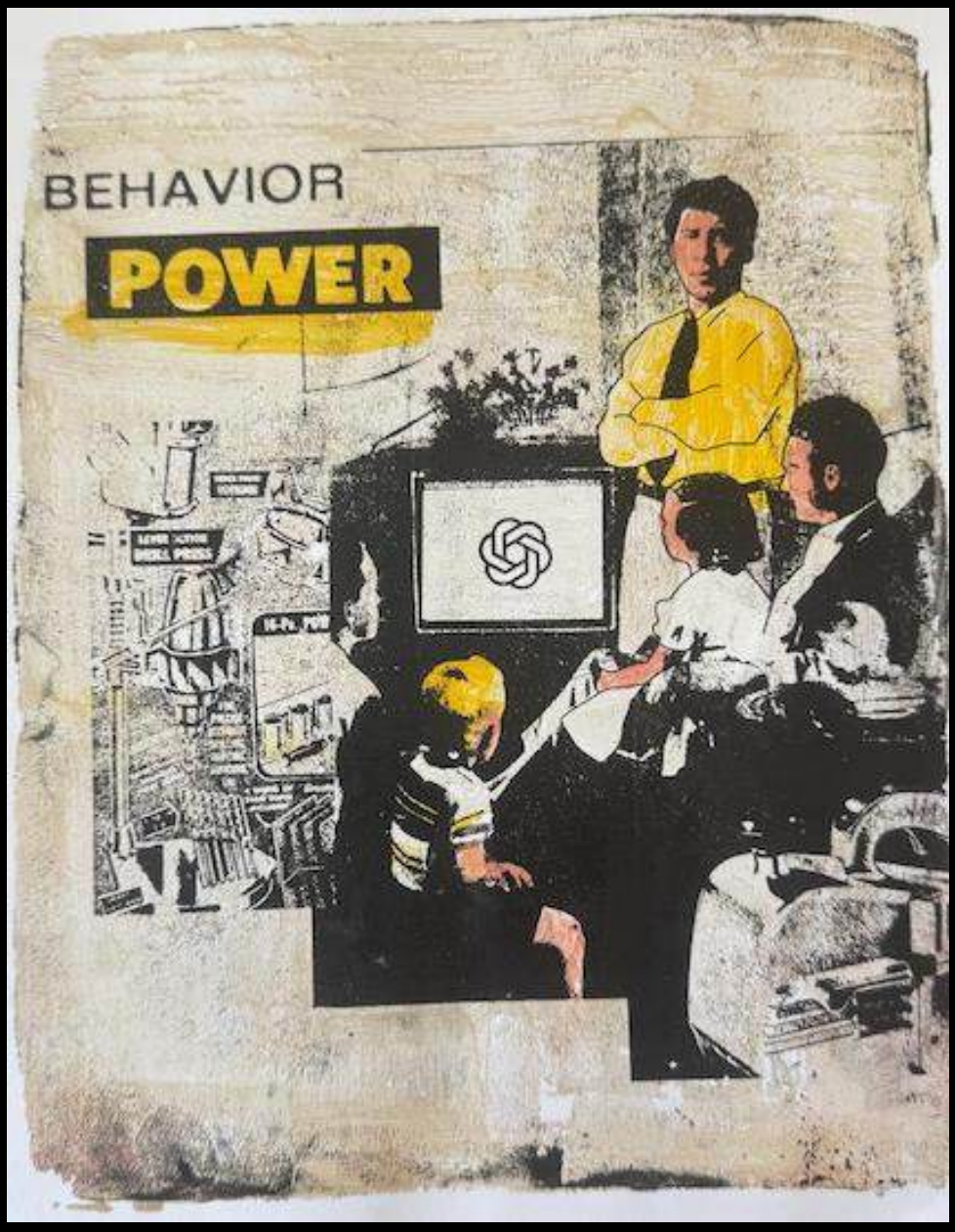
# How to explain algorithms?

**A GALLERY OF APPROACHES FROM THE ALGO→LIT PROJECT**




Algo → Lit
Algorithmic literacy for effective transparency in the eu

Funded by the European Union

## Introduction

This guide presents a collection of activities and projects aimed at explaining algorithms. It constitutes a resource for social/community and inclusion workers, researchers, advocates, activists or any professionals that need to present how algorithms are working in a way that is attractive for lay citizens. The range of examples could also inspire commissioners of algorithmic literacy projects such as State organisations, schools or data regulators. Note that our project is mainly targeted at communities impacted by algorithms in France, Belgium and the Netherlands. The selection of examples is not comprehensive but reflect a balance between many criteria:

- We aimed at presenting a diversity of formats — from serious game to workshops, from printed posters to collective forms of decision-making — coming from France, Belgium, the Netherlands, Spain, the USA and the UK,
- Considering the rise of many standardised approaches to AI literacy, we wanted to highlight examples known for the quality of their content and documentation, as well as their commitment to a critical approach to algorithms that escapes the current ethical washing,
- Leaving the hype behind, we published examples that have been overloocked, but we also includes others that are famous in academic circles but that never got the chance to be looked at by inclusion workers,
- We needed to find a balance between presenting projects that are easily reusable and the ones that might be more difficult to replicate but remain very inspiring.

## About the Algo→Lit project

The Algo→Lit project aims at improving the algorithmic literacy of digital inclusion workers to better support citizens' understanding of algorithms as well as effectively exercise their right to transparency. At a time when citizens are increasingly facing opaque algorithms, digital inclusion workers — whether they are named "helpers", "mediators", social workers dealing with the automation of public organisations or public servants in local districts — are having to deal with increasing demands about these systems. We seek to give these professionals the skills and tools they need to help people understand the systems and use their rights. The ALGO->LIT is coordinated by Datactivist, the French coop specialised in opening data and algorithms, in partnership with La Mednum (Paris, FR), the French coop representing the network of digital inclusion workers, as well as research centers FARI (Brussels, BE) and Waag FutureLab (Amsterdam, NL).

If you think we missed an example, please feel free to reach us at: algolit@datactivi.st

To know more about the Algo->Lit project, visit our website.

Funded by
the European Union

Typology of examples presented in this document:

### 1. COLLABORATIVE WORKSHOPS

Collaborative workshops are hands-on gatherings of between 1 to 3 hours dedicated to actively engaging audiences in learning about algorithms and AI effects. Most often performed by crafting a representation of the problems at stake aided by templates.

### 2. COLLECTIVE DECISION-MAKING AND SUPERVISION MECHANISMS

Collective decision-making and supervision mechanisms are initiatives where a group of citizens have an opportunity to participate in and moderately (re)shape the governance, design or effects of a given algorithmic system. They usually take the form of citizen juries, citizen's conventions/assemblies, supervision mechanisms, observatories, data donations projects or other forms of crowdsourcing initiatives.

### 3. STATIC AND PEDAGOGICAL DIAGRAMS

Static and pedagogical visualisations or diagrams are complex images that aim at giving visual representations of an algorithmic mechanism. They help to simplify and communicate difficult to see dimensions of algorithms. Moreover, they sometimes aim at giving an exhaustive picture of the many actors involved in the making of algorithms or at portraying algorithms differently from their designers, for example by emphasizing dimensions related to the social or environmental impacts of algoritms and AI.

### 4. ALGORITHM DOCUMENTATIONS AND DATASET STORIES

Documentations are written documents that explain, comment and justify the choices made during the design of a given algorithm. They can inform its outputs and effects. Dataset stories are narrative forms of storytelling aiming at documenting the data used to train or refine an AI system come from, how they were transformed, their attributes and pitfalls.

### 5. SIMULATORS AND DEMONSTRATORS

An algorithm simulator or demonstrator provides a dynamic interaction with an interface simplifying calculations, mobilizing heterogeneous data while giving the impression that one can directly manipulate an algorithm. Compared to a written algorithm or its documentation, the benefit surplus of a simulator is to test different parameters and configurations, to perform a playful interaction that promises to be a heuristic experience.

## I) Collaborative Workshops

### DESCRIBING YOUR LLM PRACTICES

Screenshopt of the projecT websiTe

The project Ecologies of LLM, developed by designers and sociologists at Sciences Po medialab, asks the following question: how can we reframe the role of Large Language Models (LLMs) like ChatGPT in ordinary work practices? For them, there's an urgent need to make room for alternative framings of AI, to move beyond predictions and look at the present. The project's team call to begin the slow and silent work of noticing what AI is already doing to our work. Only once we have paid attention to our practices can we reframe this technology on our own terms and choose how to engage with it. More importantly, observing how AI takes hold in our professional routines might bring into view the systems of labor that allowed it to happen. What's at stake is not just how we work with AI but who decides what work should look like. The projet Ecologies of LLM offers you a process, a protocol, to document closely how AI does and could affect your work. Their book acts as an artificial lens to support your inquiry. In a way, we can see their guide as a resource to reflexively investigate your own uses of LLM and share your views with a group of professionals sharing similar problems and surprises.

Below are examples of the type of exercices you could perform thanks to the book:

# Ex$^{2b}$ – All the *Things* You *Could* Do

## *Instructions*

Think about the tasks you currently do without the help of LLMs and where you could benefit from their assistance.

This list isn't about automation (delegating to the machine) but augmentation (doing more tasks or diversifying tasks). Think about activities or projects you haven't attempted yet or tasks you might hesitate to delegate.

Even if you suspect that an LLM might perform poorly on a given task, include it in your list: this exercise is about exploring potential opportunities rather than evaluating performance.

# Ex$^{3a}$ – Rough *Impressions*

## *Instructions*

1. Set a timer and spend 5 minutes reflecting on your overall LLM use. Don't look at your conversation history yet!

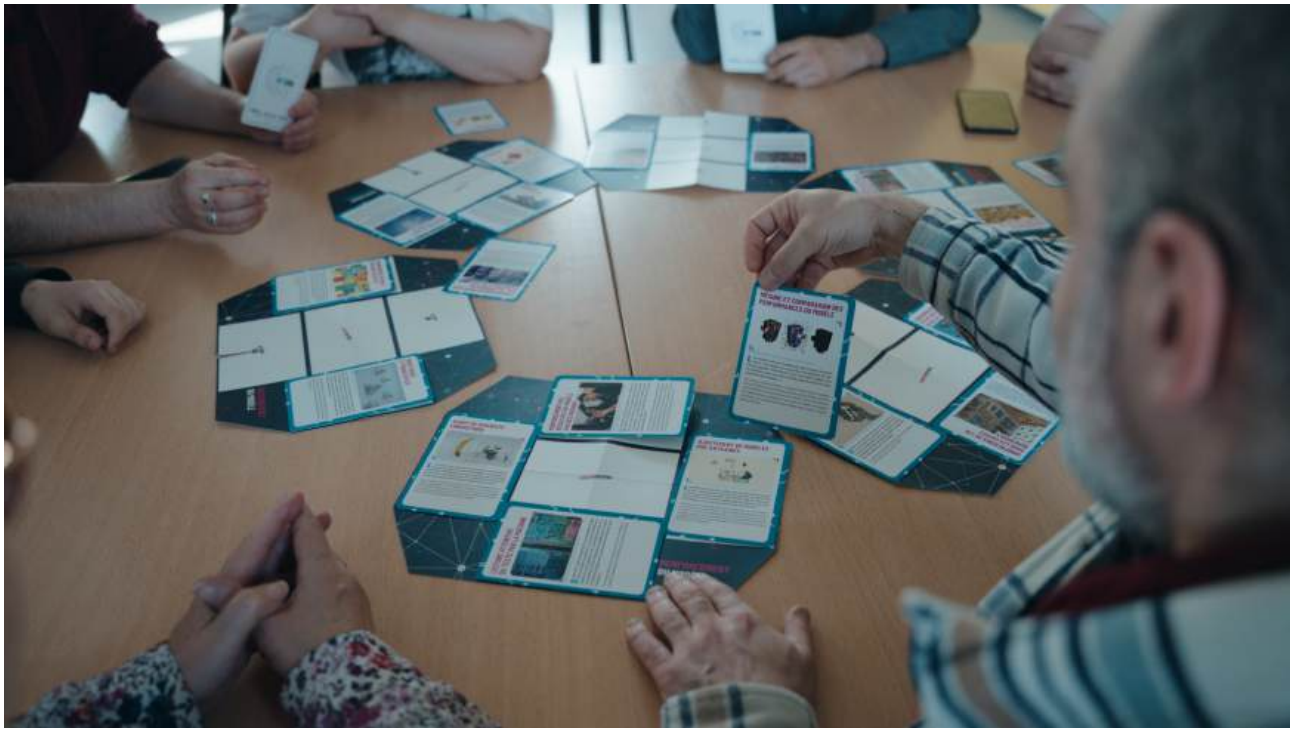
   → Assess which tasks you delegate and assign a rough percentage to each (e.g., one-third of my tasks are translation, and the rest are drafting email answers). Consider how frequently you use the LLM (daily, weekly, or less frequently).

   → Reflect on the structure of your interactions and decide how many of your conversations are: short exchanges (a single request and answer); longer conversations (several iterations); or extended dialogue (conversations that last more than a day).

2. Finally, recall some memorable moments in your use of the LLM. Think back to two or three conversations that left a lasting impression.

   → Describe them briefly and why they were memorable.

## EXPLORING GEN AI WRITING TOOLS WITH THE AI BLACK BOX



A session of the game in Nantes in April 2025

How should we position ourselves in relation to a technology that even experts struggle to keep up with? The AI Black Box is an educational game created by the cooperative Datactivist and the network of inclusion workers of Nantes Métropole (France). It is designed to help the general public understanding how text-generating AI works. Players uncover the life cycle of a generative AI product such as ChatGPT from the extraction of minerals to the energy costs of data centers. In doing so they also debate the societal, environmental and ethical consequences of AI. Constrating with many similar formats, The AI Black Box is focused on the making of AI and not simply on debating its associated impacts on work, social life or nature. The first part of the game consist in reconstructing the main steps leading to the creation of a text-generation AI (eg. training, fine-tuning, etc). In the second part, players have to place cards corresponding to a sub design step (eg. the card about filtering data has to be placed in the step dealing with data cleaning).

In this game, many computational processed are simplified without jargon. Take it as a dive into the heart of an AI system. It is best played in small groups of 7 people. The game has no prerequisites. It is a workshop open to everyone. The format is minimalist, open source (CC-BY-SA) and easily reusable. The kit includes cards you could easily print, advice on how to facilitate the workshop and a deck of slides to initiate the session.

At least I'll be able to demystify AI with a game. For me, it's still a blessing to have a game to debate about.

- *Participant of the AI Black Box*

I liked getting an overview of all the steps involved in building AI. It helped me understand the uses and environmental issues surrounding AI.

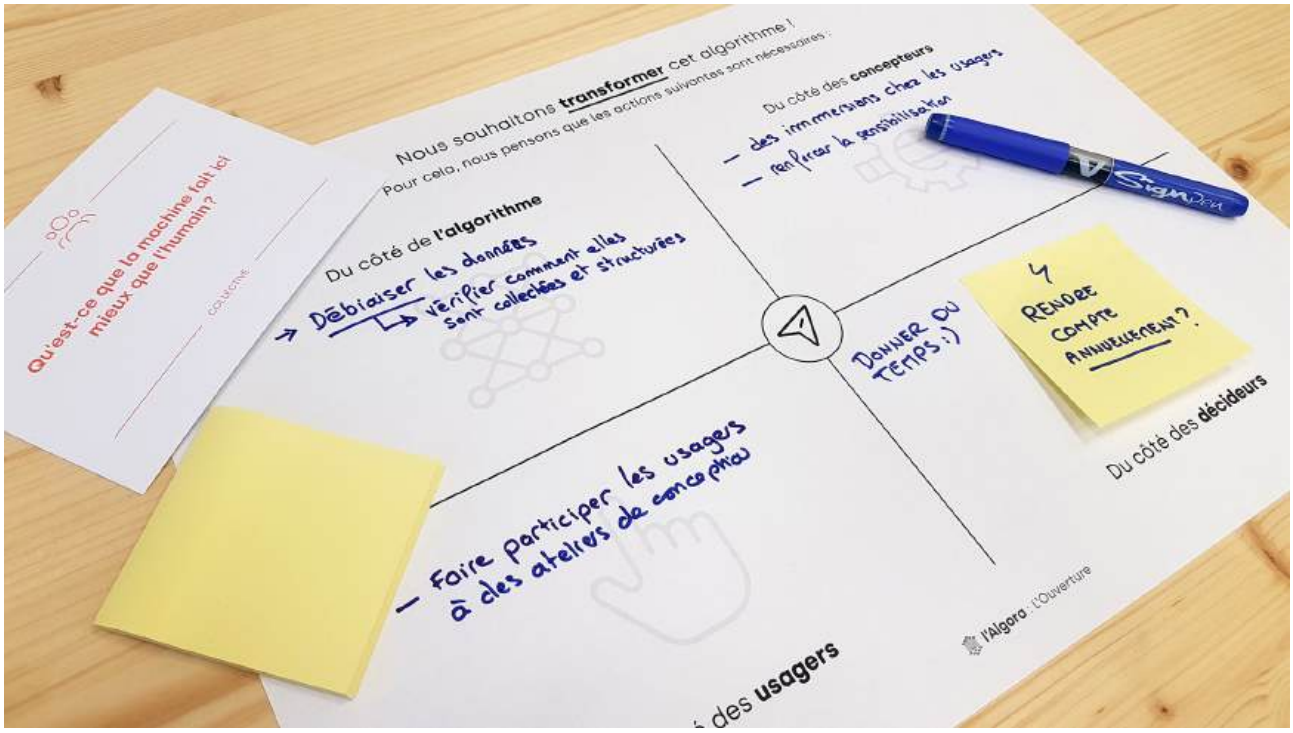
- *Participant of the AI Black Box*

I liked reorganizing the steps. I liked the social game even more, as AI becomes a social object again and not just a technical object.

- *Participant of the AI Black Box*

An example of two "action"cards

## DISCUSSING THE DESIRED IMPACTS OF AI WITH THE ALGORA




An image of the poster template used in a workshop

The Algora is a discussion tool to understand the desired, unexpected and unwelcome impacts of an existing and running algorithm, and to act upon it. It has been developed by Design Friction and commissioned by Etalab, the open data team of the French State. Designed to be used over the course of a collective workshop, the Algora facilitates communication between different profiles (decision-makers, designers, developers, civil servants, users). Through a conversation cards deck, each participant is invited to express their point of view and to confront it with the others' realities. The objective of the game is to collectively identify what the discussed algorithm actually produces. The Algora kit includes: a deck of 40 discussion cards, synthesis and action boards, resources for facilitation, and a version for online workshops.

## DISCOVERING AI WITH ST[IA]MMTISCH

The game board

The StIAmmtisch game devised by the Regional Academic Delegation for the Grand Est region (France) is inspired by the tradition of Stammtisch (a traditional table setting in Alsace and other German-speaking regions). Above all, it's about getting together to share a good time, around a topical subject: artificial intelligence. It's an easy to use board game for 2 to 6 players, or teams, accompanied by a facilitator. The game lasts around 50 minutes and is designed for classroom use. However, the game can be played by adults as well as students. Junior and intermediate levels are currently available. An expert level has been planned since the game's inception, but is not yet available.

## ORGANISING A HUMAN PARLIAMENTS ABOUT DIGITAL TECHNOLOGIES

A human parliament in 2023 taken place at Saint-Jean square near the Brussels Francophone Parliament

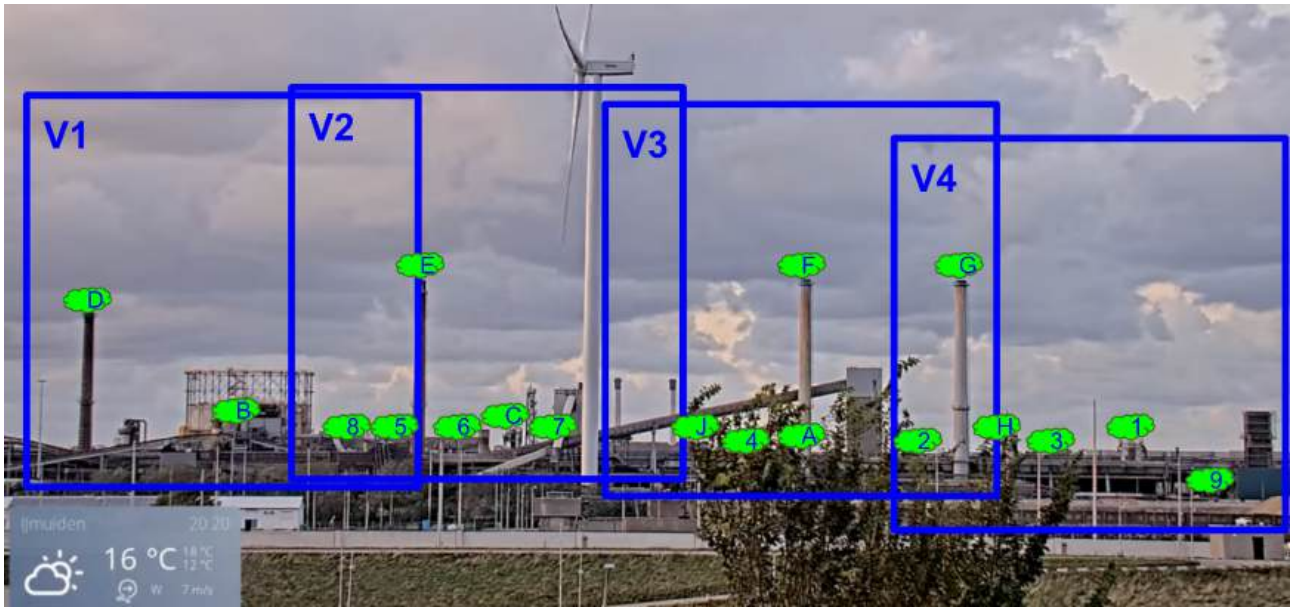
A "Human Parliament" as organised by the belgium "Comité humain du numérique" l is a time for talking and listening, sharing our experiences, our suffering and helping each other about the digital in our everyday lives. It is designed as a place where we speak from our own experience, not from theory or ideas. A space where everyone is an expert and where a fairer world can exist; a leadership tool, to build laws, collective rules or demands. This can take time and require several meetings: efficiency is not a priority. The aim is to defend the human face of digital technology in our lives during a meeting and by writing laws to ensure that digital technology adapts to citizens and not the other way round. Upon the Human Parliament, a Human Code of the Digital was published: it is a self-proclaimed citizen's law that has been in the making since 2021. This book brings together more than 3 years of actions and encounters in Brussels and Wallonia: meetings with citizens, legal researchers, politicians, discussions, workshops, cardboard set creations, shared meals, demonstrations, etc.

## ORGANISING A CITIZEN CONVENTION ABOUT THE USE OF CHATGPT

The group of students and teachers at Sciences Po Aix participating in the convention

With the acceleration of ChatGPT use by students in higher education and research, there is an urgent need to rethink its place in the pedagogical relationship and support its use in student training. To meet this challenge, the university Sciences Po Aix (France) launched, from March 6 to May 12, 2025, a University Citizens' Convention on Generative AI. This unprecedented initiative aimed to open up the conversation on the subject by creating a space for exchange and debate involving all members of the school's educational community: teachers, students and administrative staff. At the end of its work, the members of the Convention submitted to the school's management a series of proposals for training and teaching at Sciences Po Aix in the AI era.

## LABELLING DATA FOR AI WITH THE IJMONDCAM PROJECT




Images of toxic clouds as maped by the neural net

Waag FutureLab and the team of Yen-Chia Hsu at the University of Amterdam organised two meetings where residents of the Ijmond region (Netherlands) could have a say in labelling the data of an AI system aimed at detecting smoke. In other words, it aimed at accurately determining what a suspicious smoke cloud is. The AI system shows a picture with a square marker around the suspicious smoke cloud. Residents can adjust this marking to make the system more accurate. Because the community labels data to create a training dataset, people have an influence on what choices the model makes. While deep neural networks have proven useful in various applications (e.g., object recognition), training such a model requires a large amount of labeled data. Labeling all of this data yourself can easily take hundreds of hours for a single study. This is why they need volunteers. The system determines a label based on all the labels provided by researchers and citizens. Each video or image is viewed by at least two volunteers and one researcher. If the volunteers have assigned the same label to the video or image, the system saves it. Otherwise, the process repeats. The result is therefore determined by majority vote. Previous work on air quality monitoring shows that visualizing evidence of smoke emissions helps attract the attention of regulators. It also helps build community confidence in addressing air pollution.

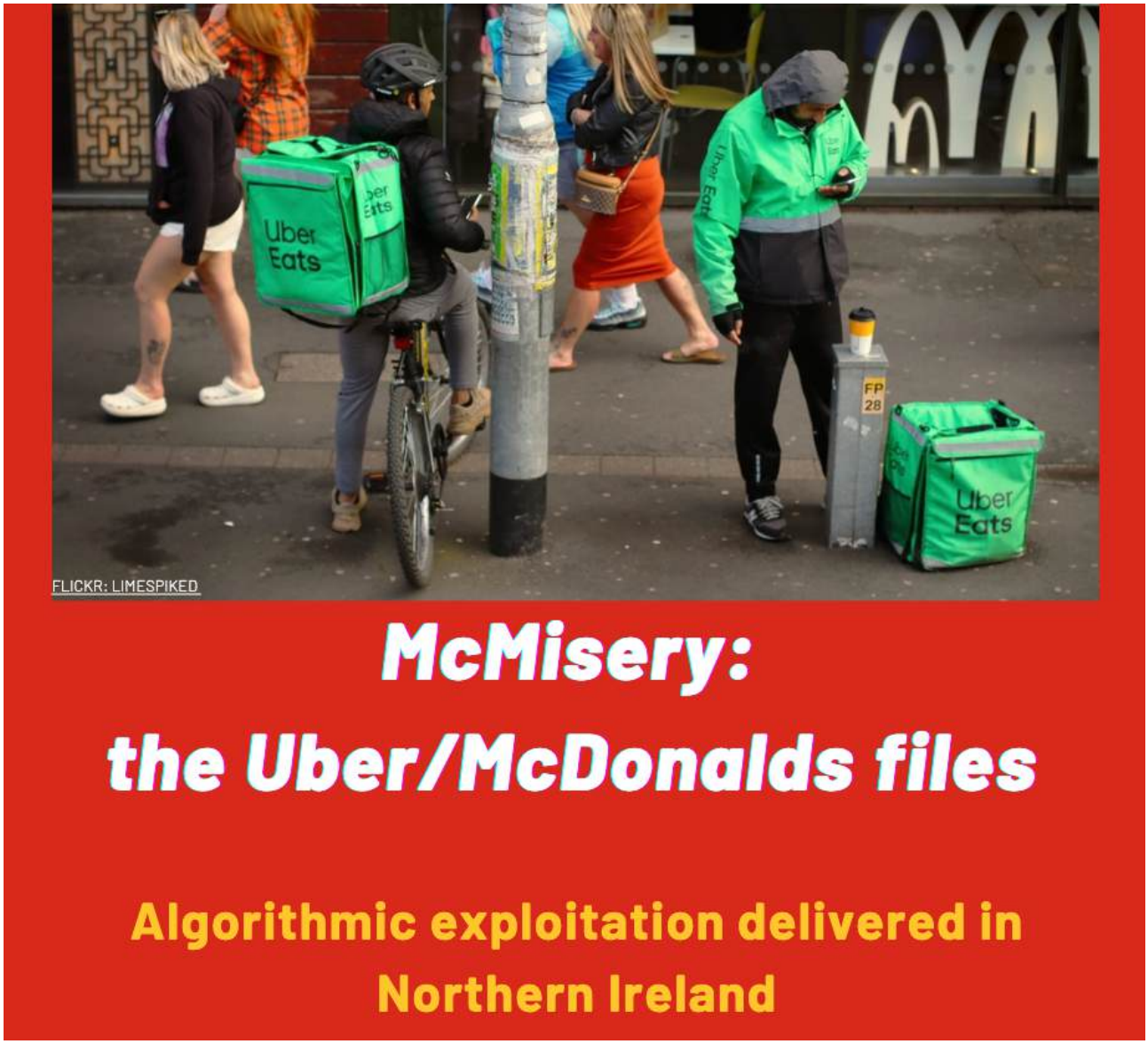
## ASSESSING THE RISKS OF AI WITH THE "LABOR AI AUTO-TEST"

Before integrating an AI system, it is essential to assess its potential benefits and risks. As an initial diagnosis, the self-assessment tool develop by the research project Labor AI (INRIA, France) allows you to gauge the critical aspects associated with integrating an AI system into an organization and identify areas for vigilance. Without prejudice to legal obligations regarding the introduction of new technologies in companies, this tool offers an initial assessment of the risks associated with introducing an AI system.

## CROWDSOURCING DATA WITH THE WORKERS' ALGORITHM OBSERVATORY

The Workers' Algorithm Observatory (WAO) is a crowdsourced auditing collaboration based in the USA and hosted at Princeton University. WAO helps workers and allies investigate black-box algorithmic systems. They crowdsource the data necessary for meaningful, scientific audits, as well as build tools and support for workers and allies. It was launched in 2022. Currently a nonprofit initiative with funding from the Mozilla Tech Fund 2023 "Auditing AI" cohort. WAO projects include for example FairFare, collective efforts to helps drivers, organizers, and policy-makers understand the ride hail industry by crowdsourcing fare data from drivers. A previous project was The Shipt Calculator another worker-led study of a black box algorithm used by the delivery company Shipt. Willy Solis, a member of the Gig Workers Collective who led the study, collaborated with Coworker.org and WAO's team to crowdsource and analyze data shared by over one hundred workers using a custom texting bot. They found that wages decreased for a majority but increased for a minority, suggesting that the algorithm determines more than just pay – it affects worker agency and well-being, too.

## HELPING WORKERS FIGHT ALGORITHMS WITH WORKERINFOEXCHANGE




Cover of the "McMisery" investigation regarding Uber drivers and Mcdonald's in Northern Ireland

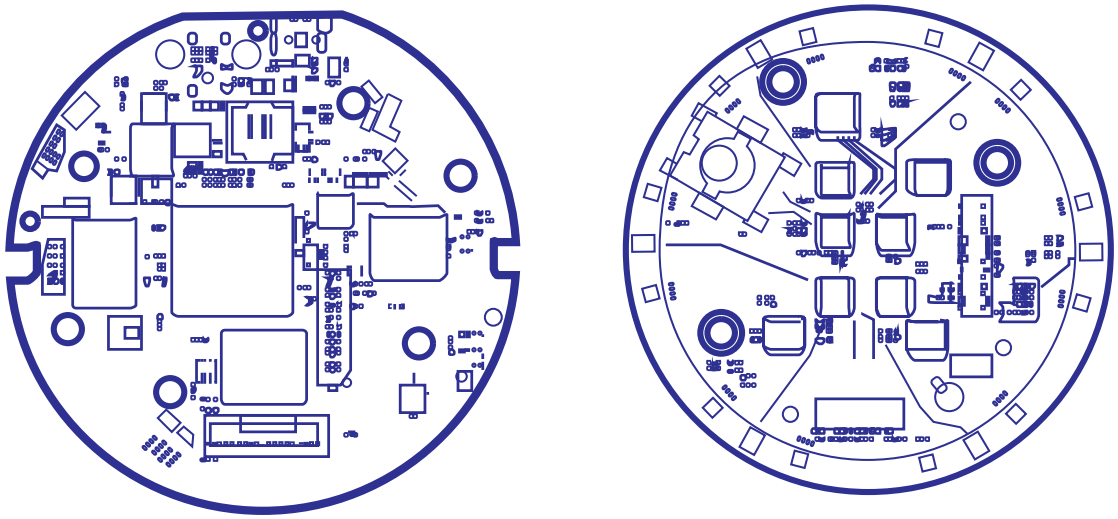
WorkerInfoExchange is a UK based data cooperative and a process to gather personal data and cases of algorithmic discrimination to reconstruct systemic issues with algorithmic management. They for example showed that Uber's introduction of dynamic pay has reduced driver pay, increased Uber's commission, and made pay more unpredictable and unequal for its workforce. Uber's so-called 'dynamic' pay system uses artificial intelligence and machine learning to set pay and assign work in real time — with no transparency for drivers and no ability to challenge automated decisions for pay and work allocation.

## BEYOND ROBOTS: WHY WE NEED BETTER IMAGES OF AI

Jamillah Knowles & Digit / Better Images of AI / CC-BY-4.0

Type “AI images” into your search engine and you will notice a pattern. Go on, give it a go! The result is striking, and it’s the same on photo libraries and content platforms. In fact, the lack of variety, and the inaccuracy is almost inescapable. The predominance of sci-fi inspired and anthropomorphised images, and the lack of readily accessible alternative images or ideas, make it hard to communicate accurately about AI. These AI images also add to the public mistrust of AI, a growing problem for innovation in a field that is sometimes seen as biased, opaque and extractive. The current dominant images reinforce dangerous misconceptions, and at best limit the public understanding of the current use and workings of AI systems, their potential and implications. We need images that more realistically portray the technology and the people behind it and point towards its strengths, weaknesses, context and applications. The Better Images of AI project aims to create a new repository of better images of AI that anyone can use, starting with a collection of inspirational images. The first stage of this project is designed to explore what these new images might look like, and to invite people from different creative, technical and other backgrounds to work together to develop better images. In creating new imagery we need to consider what makes a good stock image. Why do people use them and how? Is the image representing a particular part of the technology or is it trying to tell a wider story? What emotional response should the audience have when looking at it? Does it help people understand the technology and is it an accurate representation?

## DISCOVERING THE ANATOMY OF AN AI SYSTEM



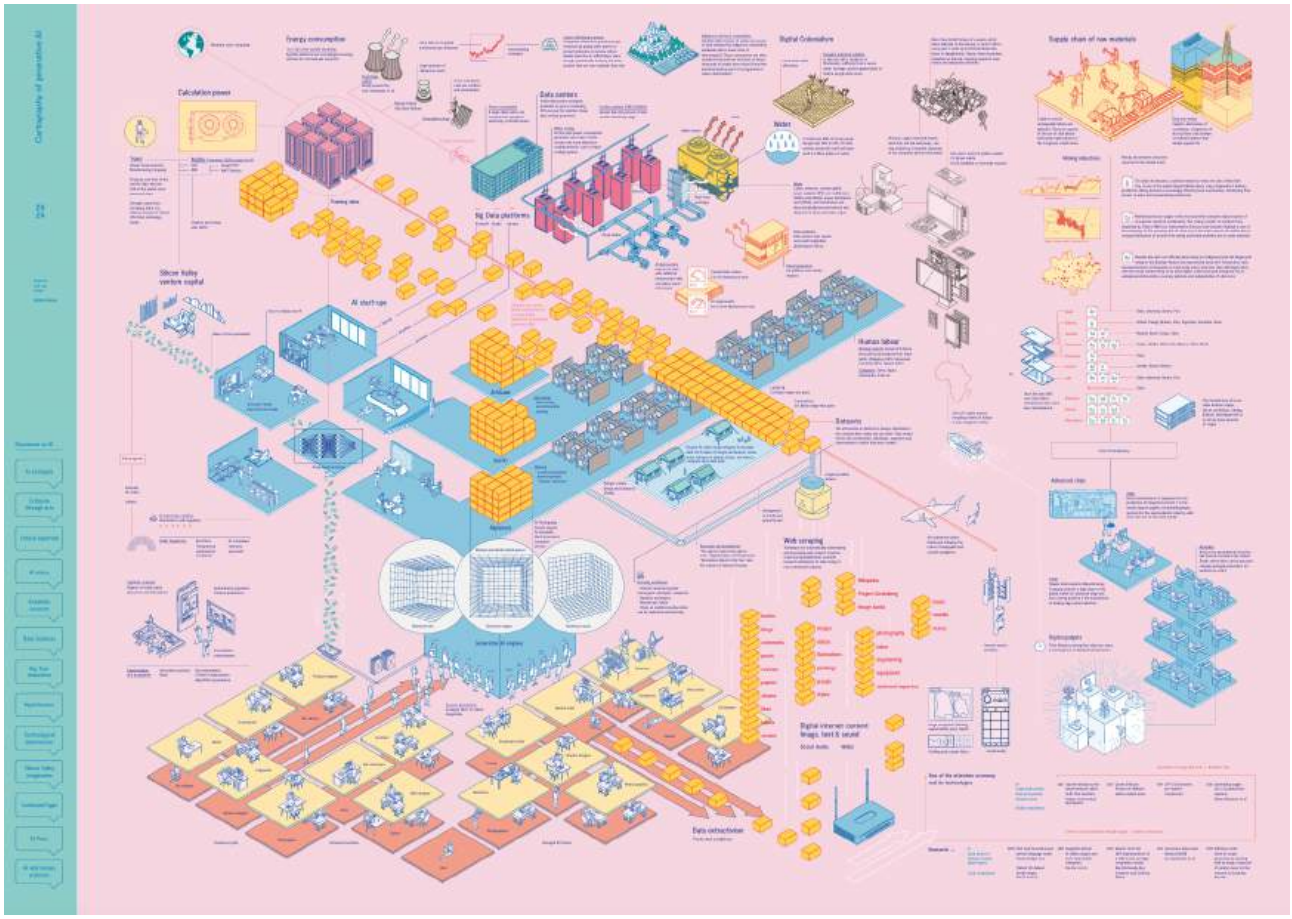
A synthetic view of the amazon echo circuit

Anatomy of an AI System is the anatomical map of human labor, data and planetary resources necessary to make the Amazon Echo device work. It was created by artist Vladan Joler and researcher Kate Crawford.

> A cylinder sits in a room. It is impassive, smooth, simple and small. It stands 14.8cm high, with a single blue-green circular light that traces around its upper rim. It is silently attending. A woman walks into the room, carrying a sleeping child in her arms, and she addresses the cylinder.
>
> 'Alexa, turn on the hall lights'
>
> The cylinder springs into life. 'OK.' The room lights up. The woman makes a faint nodding gesture, and carries the child upstairs.
>
> *– An interaction with Amazon's Echo device*

In this fleeting moment of interaction, a vast matrix of capacities is invoked: interlaced chains of resource extraction, human labor and algorithmic processing across networks of mining, logistics, distribution, prediction and optimization. The scale of this system is almost beyond human imagining. How can we begin to see it, to grasp its immensity and complexity as a connected form? This project starts with an outline: an exploded view of a planetary system across three stages of birth, life and death, accompanied by an essay in 21 parts. Together, this becomes an anatomical map of a single AI system.

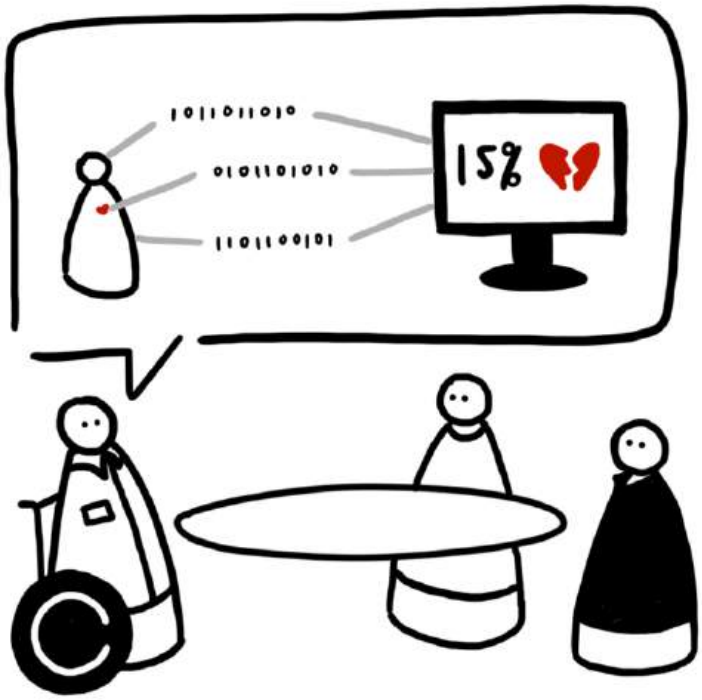
## VISUALISING THE INFRASTRUCTURE BEHIND GENERATIVE AI



The whole Cartography of Generative AI Poster

What set of extractions, agencies and resources allow us to converse online with a text-generating tool or to obtain images in a matter of seconds? The Cartography of Generative AI created by the Spanish design studio Estampa maps the whole supply chain of a Gen AI from the dataset modeling to its needs for energy, presenting at the same time many of its dimensions regarding computing power, labor forces and economy. The set of relationships presented here forms a mosaic that is difficult to grasp because it involves the linking of objects and knowledge of different kinds and scales. The discourses surrounding AI often have a strong mythical charge and are accompanied by a series of recurring metaphors and imaginaries: algorithmic agencies detached from human action or, the non-negotiable technology that imposes the future on us, the universality of data. The set of discourses that surround these technologies, whether they are more specialised or more popular, end up shaping them in one way or another. For this reason, the Cartography of Generative AI project is based on the motivation to offer a conceptual map that covers a large part of the actors and resources involved in this complex and multifaceted object we call Generative AI. Drawing on a long genealogy of critical cartographies, this visualisation aims to map the phenomenon, taking into account the tensions, controversies and ecosystems that make it possible.

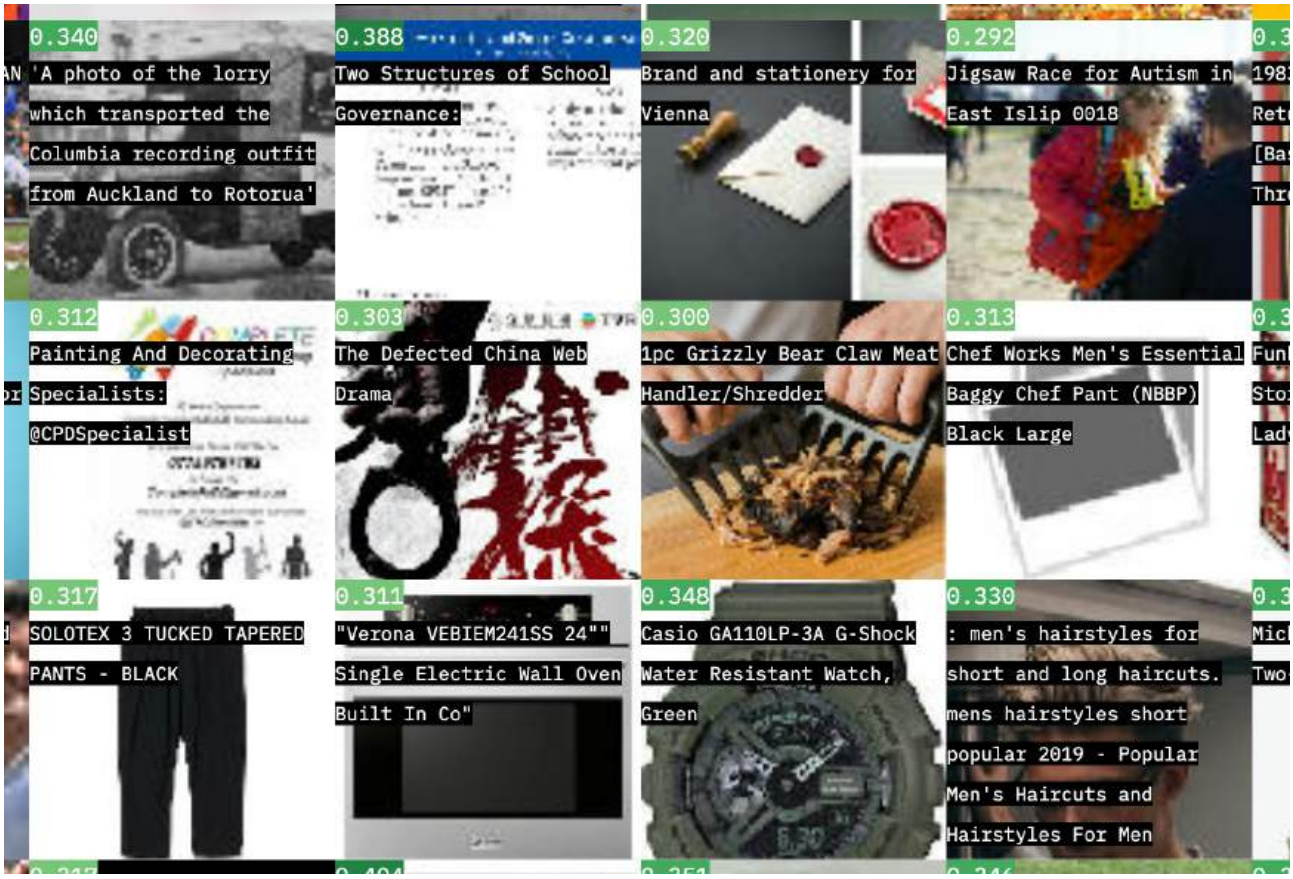
## HOW MATCHING WORKS? THE CASE OF KIDNEY TRANSPLANTS




Prediction by Yaning Wu / Better Images of AI / CCBY-4.0

The French Biodedecine Agency published a guide to understanding the score that determines kidney transplant allocation. This is taking place in a context where the shortage of transplants continues to raise difficult medical and ethical questions about the selection process for patients who will be allocated a deceased donor graft. Graft allocation is a complex and sensitive process, both in its conception and execution. There is no global or definitive solution. It varies from country to country, depending on the allocation criteria taken into account, and over time, depending on the periodic evaluation of results. An allocation system must be adapted, efficient and as equitable as possible to the health needs of the patients concerned. For some patients, one solution is to offer transplants on a “priority” basis. This is the case for patients in emergency situations or with hyperimmunity, who are currently given priority, generally at national level. For the vast majority, it is impossible to prioritize one group over another. The allocation system must therefore achieve a compromise between fairness, efficiency and feasibility, and take into account several allocation criteria at once. One effective solution is to use a score. This official documentation from the French Biomedical Agency explains how this score is calculated.

## WHAT ARE THE DATA BEHIND AI?




A screenshot of the story showing images metadata

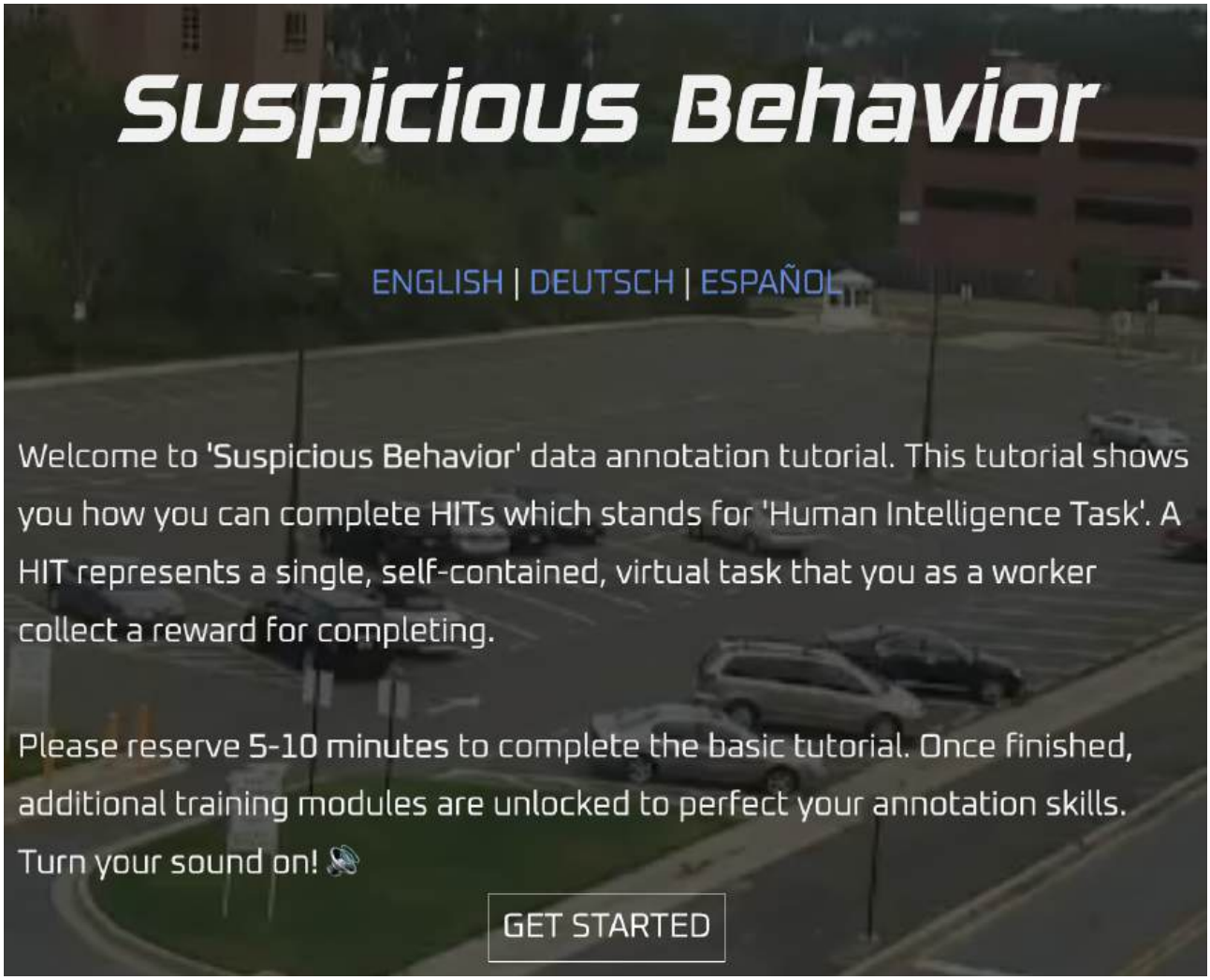
Models all the way down is a visual story that can be understood as a narrative form of documentation. It tells the story of LAION-5B, an open-source foundation dataset used to train AI models such as Stable Diffusion. It contains 5.8 billion image and text pairs —a size too large to make sense of. In this visual investigation created by Christo Buschek and Jer Thorpthe from the Knowing Machines research project, we follow the construction of the dataset to better understand its contents, implications and entanglements.

## DEMONSTRATING THE RELIABILITY OF NEURAL NETWORKS WITH PYRAT

The PyRAT demonstrator

A team of researchers at CEA-List, a research lab part of France's public technology and energy research organization, has been developing a new software named PyRAT. Designed to prove the reliability of neural networks' response in given configurations, PyRAT can help assess the trustworthiness of AI-driven systems in critical applications ranging from aircraft collision avoidance to medical diagnostics. The laboratory was eager to demonstrate PyRAT to a wide audience, including researchers, scientists, industrial partners or prospects in a way that is both technical and educational, through a demonstrator presented in the CEA's showroom in Paris-Saclay. The French design studio Units worked with the laboratory on how to best represent and explain the highly complex and opaque processes at work in large neural networks. The result is a demonstrator uncovering the logic behind PyRAT and the various mathematical operations at play, through which users can experiment with different analysis parameters. The interface is on display inside the CEA's showrooms where it can be controlled through a dedicated hardware controller that Units designed and developed.

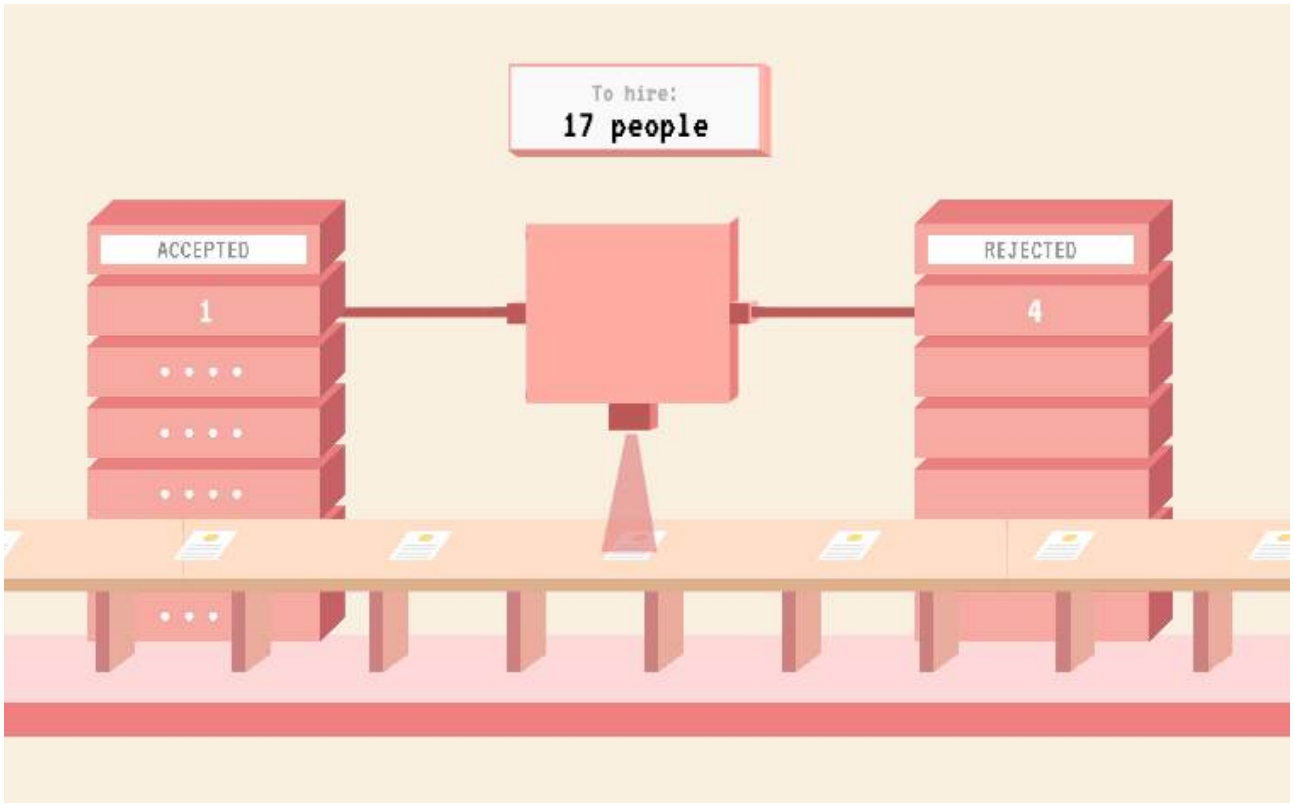
## HOW ALGORITHMS INTERPRET IMAGES OF SUSPICIOUS BEHAVIORS?




A screenshot of the game

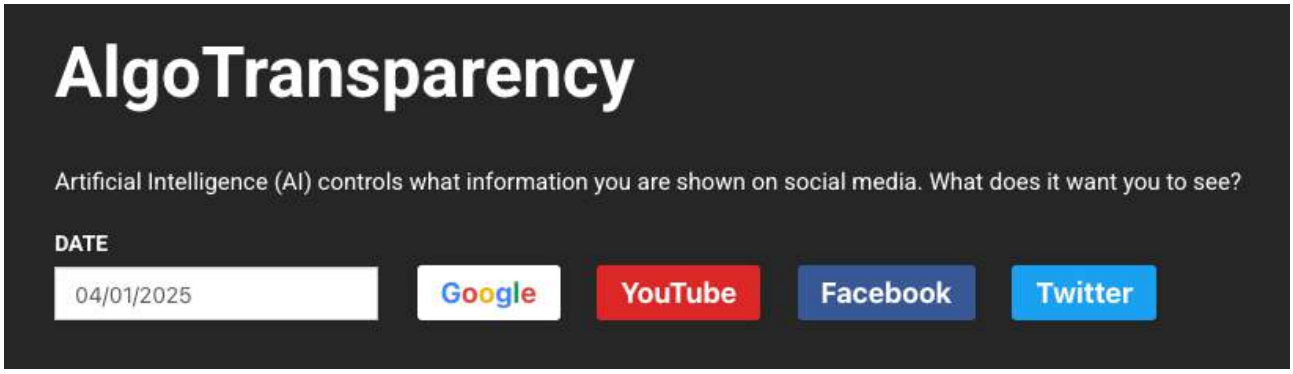
The online game Suspicious Behavior developed by the artists Kairus shows a world of hidden human labour, which builds the foundation of how 'intelligent' computer vision systems interpret our actions. Through a physical home office set-up and an image labelling tutorial the user experiences the tedious work of outsourced annotators. In an interactive tutorial for a fictional company the user is motivated and instructed to take on the task of labelling suspicious behaviour.

## LEARNING ABOUT THE BIAS IN HIRING ALGORITHMS




Survival of the Best Fit is an educational game about hiring bias in AI. It explains how the misuse of AI can make machines inherit human biases and further inequality. Much of the public debate on AI has presented it as a threat imposed on us, rather than one that we have agency over. Designers of the game wanted to change that by helping people understand the technology, and demand more accountability from those building increasingly pervasive software systems. The project was developed by Gabor Csapo, Jihyun Kim, Miha Klasinc, and Alia ElKattan, and supported by the Creative Media Award from Mozilla Foundation. With this game they wanted to reach an audience that may not be the makers of the very technology that impact them everyday. it is important to make people understand how AI works and how it may affect them so that they can better demand transparency and accountability in systems that make more and more decisions for us.

## HOW YOUTUBE'S RECOMMENDATION ALGORITHM WORKS?




2 billion people use YouTube monthly. YouTube's recommendation algorithm determines what people watch for more than 70% of the views. That's 700 million hours - or 1,000 human lifetimes - every single day. The project AlgoTransparency monitor 800+ top information channels, including news channels, vlogers, and shows from across the political spectrum.

Insert text

A guide created by the Algo->Lit project, funded by the Erasmus + Agency, coordinated by Datactivist, and developed with the organisations Waag Futurelab, Fari and Mednum.
Credit of the cover image: Bart Fish & Power Tools of AI / https://betterimagesofai.org / https://creativecommons.org/licenses/by/4.0/
First Published in September 2025.
Published through a [CC BY-NC 4.0 licence](https://creativecommons.org/licenses/by-nc/4.0/deed.fr).

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the French Erasmus + Agency. Neither the European Union nor the granting authority can be held responsible for them.

For any questions, please feel free to reach us at:
algolit@datactivi.st
To know more about the Algo->Lit project, visit our [website](https://www.algolit.eu/).

Funded by the European Union