

DOI: 10.35643/Info.26.1.2

Dossier temático: Ética de la Información

Dread Data and Algorithms

Datos que intimidan y algoritmos

Dados e algoritmos de medo

Stefan Ullrich^a

^a Weizenbaum Institute for the Networked Society, Research Group Lead, Germany, Hardenbergstr. 32, 10823 Berlin. ORCID: [0000-0003-2541-0412](https://orcid.org/0000-0003-2541-0412). Correo electrónico: stefan.ullrich@tu-berlin.de

Abstract

The mystical aura surrounding algorithms we observed in the last years is currently more and more perceived as an annoyance that we must accept but not like. Nevertheless, algorithms are still praised as a panacea or are seen as the main culprit of «digital immaturity» (digitale Unmündigkeit). The current debate is less about the algorithms (as a mathematical entity) themselves than about their technical implementation in hardware and software, often used synonymously with «Artificial Intelligence». However, the enormous informational transformation potential of algorithms and data is already visible in the mathematical construct as we will see.

Keywords: ALGORITHMS; DATA LITERACY; TURING GALAXY; TECHNE (τέχνη).

Resumen

El aura mística que rodea a los algoritmos, observado en los últimos años, se percibe cada vez más como una molestia que debemos aceptar aunque no nos guste. Los algoritmos tanto siguen siendo elogiados como una panacea, como considerados culpables de la «inmadurez digital» (digitale Unmündigkeit). La discusión en curso sobre los algoritmos se refiere menos a su entidad matemática que a su implementación en tecnología y programas informáticos, a veces considerada como sinónimo de «inteligencia artificial». Sin embargo, el enorme potencial de transformación informativa de los algoritmos y los datos ya es visible en la construcción matemática, como podrá apreciarse en este trabajo.

Palabras clave: ALGORITMOS; ALFABETIZACIÓN DE DATOS; GALAXIA TURING; TECHNE (τέχνη).

Resumo

A aura mística que envolve os algoritmos que observamos nos últimos anos é atualmente cada vez mais percebida como um incômodo que devemos aceitar, mas não gostar. No entanto, os algoritmos ainda são elogiados como uma panacéia, ou são vistos como culpados pela "imaturidade digital" (digitale Unmündigkeit). O debate atual sobre a ética de dados e algoritmos, se refere menos à entidade matemática, do que à sua aplicação técnica, às vezes, especialmente à "inteligência artificial". No entanto, o enorme potencial de transformação informacional dos algoritmos e dados já é visível na construção matemática, como veremos.

Palavras-chave: ALGORITMOS DESMISTIFICADOS; ALFABETIZAÇÃO DE DADOS; ÉTICA DA INFORMAÇÃO; GALÁXIA DE TURING; TECHNE (τέχνη).

Received: 20/08/2020

Accepted: 02/05/2021

1. Introduction

Many things dread and wonderful,
none though more dread than mankind

Muchas cosas hay intimidantes y maravillosas,
pero ninguna tan intimidante como la humanidad.
Sophocles, Antigone, Ode I (Chorus of the Elders)

The mystical aura surrounding algorithms we observed in the last years (Ullrich 2019a) has given way to a weariness or annoyance to the point of rage against algorithms and algorithmic decisions. British protesters in the summer of 2020 have expressed this most clearly when they chanted: «fuck the algorithm». The algorithm they referred to should have been used during the Covid-19 pandemic in the following way: Since it was not possible to take final exams, the Ministry of Education of Great Britain planned that the teachers should assess the situation and basically predict the results. To correct for a personal bias, an algorithm was applied that took into account the average grades of previous years. It quickly showed the principal flaw of this approach that this cements the status quo. Pupils

with lower grades cannot improve themselves through learning and increased performance, or only with difficulty, if the average grade of their predecessors was low. This algorithm was designed to correct human prediction by making another prediction, but the result should not be called *prediction* but rather *retrodiction* because all the data are from the past. The protests showed also that this algorithmic assessment correlated heavily with the socio-economic status and addresses of the pupils and hence revealing the discriminatory potential of seemingly neutral algorithms (cf. Orwat, 2020).

It was also algorithms that were the subject of a 2018's US-Senate hearing when Facebook chief executive officer appeared before the Senate's Commerce and Judiciary committees to discuss data privacy and disinformation. It is therefore appropriate to look at the phenomenon called «algorithms» and asking, what *is* an algorithm that it evokes such stark resistance? The current debate is less about the algorithms (as mathematical entities) themselves than about their technical implementation in hardware and software, often used synonymously with «Artificial Intelligence». In a way, this is a variation on how technology influences human action and thinking. Questions about the influence of technology on human behaviour (πρᾶξις) and the scope of knowledge (ἐπιστήμη) have been asked since antiquity, but it was only since the industrial revolution that it became clear that these are not theoretical questions, but rather concrete challenges posed by technology that can be experienced. It was not the theoretical description of energy conversion, but its technical implementation in the ever-turning steam engine that transformed society. However, the enormous informational transformation potential of algorithms is already visible in the mathematical construct as we will see. Perhaps we should start from scratch. An algorithm is a coded step-by-step instruction for a given problem to be processed by any processing unit. We all applied the Euclidean Algorithm in school to determine the Greatest Common Divisor, one of the oldest algorithms (ca. 300 BC) we still teach today. You can easily find a version online. But if you looked it up and did not pick up a pen and a paper and started drawing lines then you would just *read* the text of the algorithm not *processed* it. An algorithm written for a human being cannot force them to process it (as many schoolteachers will tell

you) but an algorithm written for a computer will do exactly this. Algorithms reveal an affordance to process them in order to get information that was not accessible before processing. The Minotaur cannot locate the exit of Daedalus' maze by reading an algorithm (even if he was literate which I seriously doubt), instead he has to process it by following a simple instruction: «Keep your right hand at a wall and follow any branch on that side. If you hit a wall with your horns turn left. Repeat until you find the exit.» Then he will eventually find his way out. «Where is the exit?» and «How do I find the exit?» are fundamentally different questions. The power of the algorithms is nowadays most evident thanks to the availability of huge amounts of data («Big Data»). Any kind of consideration of algorithms must therefore sooner or later turn to the data. In this article I would like to show, through the fundamental explanation of how data and algorithms work, that a general understanding is necessary to be able to live self-determined and empowered in the digitalised society.

2. Data

Data are coded measurements, discrete representations of a continuous world. These representations are no longer subject to the laws of nature but instead «given» to a calculating mind (like the Latin data etymologically suggests). We can talk about 14.12 pupils in primary education per teacher in Europe—and that shows the power of data: the incommensurable individual now becomes measurable, dividable, comparable. The challenge now is not to be tempted by calculation alone. Is that number, 14.12 pupils per teacher in Europe, better than 14.43 pupils per teacher in Northern America or 11.02 in Uruguay (UNESCO, 2020)? Data give us a hint where to look but ultimately, we must think about what that data mean. If you set a goal based on these data alone, such as aiming for as few pupils per teacher as possible, then this goal can also be achieved by not sending any kids to schools.

2.1 Power of Data

The General Data Protection Regulation (GDPR) of the European Union (EU) applies to the processing of personal data within the EU. To be a bit more precise here: The GDPR applies if the data subjects or the data processors or the data

controllers are based in the EU, but not if this data is purely based on and used for personal activities and without professional or commercial intent. Like other data protection and privacy laws, e.g. Chile's *Ley 19.628 sobre protección de la vida privada* or the *Bundesdatenschutzgesetz* of Germany, it focuses on the concept of personal data as opposed to non-personal data. But there is no insignificant *datum* under the conditions of automated data processing.

Let's try a small exercise to illustrate this point: Imagine that you assign each person of a group of 40 people a playing card of a Spanish deck of cards (*baraja española*), i.e. «7 de copas», «Sota de espadas», «As de oros» and so on. The request «Take the Jack of Swords (*sota de espadas*)» concerns exactly one card. You would probably agree that this is a personal datum because we have assigned exactly one person to each card in the preparation. But how different is the instruction: «Select all cards with weapons (*armas*). Put all the cards aside that show only items, also cards with horses or crowns, and finally all the clubs (*bastos*). You now have only one card left in the end, without any reference to a person—although strictly speaking you have a group of cards left, which in this case only consists of one card. Furthermore, in the second request the terms «swords/espadas» and «jack/sota» do not appear at all. It can now be argued that the second request is merely a different formulation of the first. But for the person who follows this instruction, this is only true if the person knows how many *oros*, *copas*, *espadas*, *bastos*, numbers and pictures there are in this specific deck of cards and what the distribution looks like, etc. The takeaway of this thought experiment should be: If an organisation has informational power it can discriminate individuals based on anonymous or statistical information (Pohle 2016, p.17), there is no need for the concept of personal data. The existence of processable data alone is sufficient. The separation of these two types of data is set out in jurisdiction, but I would like to argue that this separation makes no technical or moral sense. Any protection of personal rights must concern all data that is processable, not just personal data. But what does processable mean? To answer this, let's go back to the beginnings of automated data processing. Marginal punched cards are punched cards suitable for manual processing (see Bourne, 1963, p.80). An unencoded card has two rows of holes all around the

edge. Now a key, a code, is designed and the cards are notched so that a slit is created at certain points. The encoded marginal punched cards are now placed on a stack and placed so that the holes are on top of each other. If you now insert a knitting needle and lift the stack of cards, those cards with a slit at the point of the needle will fall down. The time saved when searching is enormous compared to normal index cards; the selection speed is between 30,000 and 40,000 cards per hour (Kiermeier und Renner 1960, p.317).

Now take our playing cards from our little exercise above and add in your mind corresponding slits on the edge. But unlike above, you will not be asked to select certain cards, but you simply receive knitting needles in a certain position representing a search criterion («selector»)—and in no time you will receive the result of a massive parallel operation and the «sota de espadas» is picked. This process therefore requires two stages. First, the data must be written on cards (coded data), then it must be made automatically processable by adding the slits (coded algorithm). It also shows the limitations of manual data processing: If you want to be able to select any of the 40 cards in one pass, you need six needles. If you want to select one card out of 3.5 million cards in one pass, you need 22 very, very long needles. As Wolfgang Coy (1994, p.9) fittingly describes, the real «explosive power of information processing» emerges only with the use of the electrical computers.

2.2 Computational Power

The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. It can follow analysis; but it has no power of anticipating any analytical relations or truths. Its province is to assist us in making available what we are already acquainted with. (Lovelace 1842, p. 44)

Ada Lovelace, the then unknown author of a commentary on a lecture given by Charles Babbage on his Analytical Engine, is considered the world's first programmer. In her «Notes», she wrote the first algorithm for a machine, a computer programme in a modern sense. She also recognised that the mechanical processing of numerical representations is a symbol transformation that works even if the transforming machine does not «know» what a «7» «means». Sadly, her Notes received little attention and also Babbage's contributions to science,

technology and politics—once European bestsellers—are known today only in specialist circles. It was Alan Turing (about a hundred years later) who transferred the mathematical concept of calculability to a purely mechanical transformation of symbols. His «paper machine» consisted of an infinite, graph paper, an eternally writing pen and a reading head. It was in a certain configuration, which changed according to the symbol read and the corresponding allocation table. His machine provided the definition of what is calculable: Calculable is what can be calculated by a *Turing Machine* (Turing 1937, pp.230-265).

Alan Turing was also able to put his ideas into very practical use. In the British Bletchley Park one of the first universal machines was built during World War II, not least to crack the encryption of the German Enigma. The advantages of such a universal algorithm machine are obvious: the computer does not sleep, does not go on strike and is not poached by the enemy or the competition. The algorithmisation of life, universe, and the rest has shaped our epoch so fundamentally that it should be called «Turing Galaxy» in reference to Marshall McLuhan's «Gutenberg Galaxy». When we speak of McLuhan's «Gutenberg Galaxy», we emphasise the role of the book as a leading medium («Leitmedium») for our culture. The algorithmisation of social life has had an equally fundamental impact on our epoch, which is why I use «Turing Galaxy», coined by Wolfgang Coy (1994). Already during the Enlightenment there were first thoughts about this predictability and computability of society. What Leibniz wrote (perhaps with tongue in cheek), now seems to be coming true: we simply calculate political decisions and in case of disagreement we simply calculate who is right. *Calculemus!* In the case of epistemological ambiguities, people of the Turing Galaxy that tried to grasp the essence of a problem are no wiser now than when they began their studies. Based on the reporting of research projects, it is to be feared that the guiding principle of our epoch is: What we have no idea of, we let algorithms run over it to understand whatever holds the world's innermost core together. At the same time people of the Turing Galaxy are no longer interested in the inner workings of universal algorithmic machines as long as the results produce exciting numbers (or colourful charts). This is a fatal development as Joseph Weizenbaum wrote in *Computer Power and Human Reason*:

Our society's growing reliance on computer systems that were initially intended to “help” people make analyses and decisions, but which have long since surpassed the understanding of their users and become indispensable to them is a very serious development. It has two important consequences. First, decisions are made with the aid of, and sometimes entirely by, computers whose programs no one any longer knows explicitly or understands. Hence no one can know the criteria or rules on which such decisions are based. Second, the rules and criteria that are embodied in such computer systems become immune to change, because, in the absence of detailed understanding of the inner workings of a computer system, any substantial modification of it is very likely to render the whole system inoperable and possibly unrestorable. Such computer systems can therefore only grow. And their growth and the increasing reliance placed on them is then accompanied by an increasing legitimation of their “knowledge base”. (Weizenbaum 1978, pp.236-237)

Although the fundamental objection is still valid, access to computing power has changed dramatically. In the 1980s IBM’s Personal Computer and Apple’s Macintosh were introduced targeting household tables instead of working desks. Ten years later, the previously academic Internet opened its servers with the World Wide Web and the maxim: Join in, put up some data! Ten years later, Wikipedia showed what an online community can achieve when it works together to co-design this online world we are now more and more living in. Another ten years later, we seem to be subjected to the digital world that seems to shape us more than we in return are shaping it despite the new opportunities provided by low-barrier technology. In order to keep our informational sovereignty or to gain it in the first place, we have to focus on the enculturation process through which we learn about the digital culture we live in: «The power of calculation might become a source of liberation by enculturating algorithms instead of stretching human mental creativity on the rack of algorithmically controlled computers.» (Capurro 2019, p.136)

This creative-reflexive process of enculturating algorithms was an essential part of the early years of the computer age, simply because both computing power and data availability were scarce goods. You had to understand a problem deeply in order to develop an algorithm that can point in the vague direction of a solution—for the most important problems in life are not calculable. Love, friendship, dreams cannot be computed.

2.3 Revenge of the Technai

Computers have been around for about 80 years now (plus or minus a few years, depending on who you ask). In the beginning, information processing was mainly used to speed up calculations, which were mostly of a military nature. With the introduction of database systems in the 1970s, information processing changed from calculating numbers to processing data. «Big Data», «Artificial Intelligence» and «Machine Learning» are considered as key technologies in the «Data Age» of the Turing Galaxy. A central characteristic of this latest wave of digitisation is the increased digital recombination of data with the help of algorithms and heuristics. Heuristic systems are the soulless siblings of algorithmic systems in the way that heuristic systems do not care about causality or comprehensive data, but about correlation and plausible data. The decisive dichotomy is no longer «true»/«false», but «good enough»/«unsellable». The greatest strength of heuristic systems is that these systems can also handle inaccurate, missing or contradictory data.

This poses a particular challenge for ethical considerations of algorithms. We need to understand not only the moral principles affected by technology, but also the underlying technology itself. For thousands of years philosophers have been dealing with the questions of the moral application of technical products, but not with technology itself and the way of thinking behind it. Perhaps this is the late revenge of the once so little noticed Technai. In ancient times, «technai» were understood to mean various practices of craftsmanship associated with the word *technê*. For Plato these included (following Perry 2008): medicine, horsemanship, huntsmanship, oxherding, farming, calculation, geometry, generalship, piloting a ship, chariot-driving, political craft, prophecy, music, lyre-playing, flute-playing, painting, sculpture, housebuilding, shipbuilding, carpentry, weaving, pottery, smithing, and cookery. Medicine is a «*iatrikê technê*» and the practitioner is an «*iatros*»—and not a technician. The ancient (Western) world does not speak of the technicians in this abstract form. The practitioners were considered to be bound in the world of the necessities and thus outside the political sphere and therefore they were not discussed further in the very select circles of the Platonic *Akademia*. The technicians and their works are thus hidden, unfolding their full effect without the

critical eyes of the philosophers or at least the public. However, hidden or not, technology codifies, intentionally or not, values and assumptions of the creators and the society they live in. The behaviour of individuals or the masses can be controlled or at least influenced by an appropriate technical setting. The famous example of sociotechnology is the heavy pendant of a hotel key, which one likes to leave at the reception in order to not bulge the jacket (Latour 1990, p. 104). The motives for concealing—or revealing—technology can be of very different nature. This is an important aspect for our main topic because data and algorithms are invisible and even their substrates are also becoming more and more invisible to people other than technicians. In order to use algorithms on data to get information, an appropriate technê is needed in the double sense of artefact and capability.

Important algorithms, like Google's PageRank Algorithm, are patented and therefore publicly visible. The PageRank value P of any webpage u in regard to each webpage v with L hyperlinks in the set B containing all webpages to page u is: $P(u) = \sum(P(v)/L)$. This little example shows that transparency of data and algorithms does not help if you lack the capabilities of actually using them. Without Google's data you cannot apply the PageRank algorithms meaningfully. And even if you have all the data you still need computational power. Freely adapted from Kant's First Critique: Algorithms without data are empty, data without algorithms are blind.

2.4 Data-driven Decisions

Computer systems (including early forms of machine learning systems) have been around for over 70 years. In Turing Galaxy's early years, computers were mainly used to speed up calculations (see the paragraph about punched cards above), mostly for military use (calculating trajectories or code breaking). With the introduction of database systems in the 1970s, the field of application has changed and is still changing more and more from merely calculating numbers to processing data. A central characteristic of the latest wave of digitisation is the increased digital recombination of data and the ability to process huge amounts of data («Big Data»), which can lead to power asymmetries in favour of the data

possessor. Data-driven digitisation enables a fast, simple and relatively inexpensive way to extract most varied data from most varied sources. And thanks to very low costs for storage systems these data are kept in stock. Data about citizens or customers are of particular interest because they can be used to predict or even influence behaviour (buying products or voting for a candidate). New statements about individuals and groups become possible thanks to algorithms and data. It is understandable that data play an increasingly important role in decision-making processes. Although decisions can only be made by people—a computer system calculates—it is necessary to consider the entire socio-technical system of data-based or even data-driven decisions.

Data analysis is an interaction game (in a Wittgensteinian sense) between humans and machines using interfaces. For some tasks in the data analysis process humans are better suited, for other parts the sheer computational power of modern systems surpasses human capabilities. A system designed to help on decisions is a socio-technical data analysis system. It consists of at least two subsystems: The computer system and the operator (embedded in an organisation). At the end of this complex process of assisted decision making the result calculated deep inside is just that: a calculation, not a decision. It is a number that can be used for a decision (say number of publications when deciding to give someone a permanent position in academia). This number provides security, safety and control. It suggests a certainty that no computer system can provide. Yet «it is not the computing power that makes computers seem so powerful, it is the elegance of modelling that appeals to the overworked mind.» (Ullrich 2019b, p.23, own translation).

The fundamental problem with data-driven decisions can be illustrated by a simple example. When we look at Georges Seurat's famous painting «un dimanche après-midi à l'Île de la Grande Jatte» and ask ourselves whether a particular brushstroke is part of the tree or the *Seine*, our mind decides based on position, colour, and intuition. Pointillism wanted to show exactly this: Our brain draws a virtual line where there is none in the physical world. There are no lines in nature, we draw some in nature's representation in our minds.

We also draw such lines in software systems that are supposed to classify something. At a given point, the software classifies a *pixel* in question as a tree or as part of the Seine. Perhaps a sensor measures the wavelength of the colour and prints out a number, say, 490 nanometres. Is that still blue? Or already green? Instead of the result blue or green, the software should actually print out the confidence interval. «Can be blue, can be green.» How does the system «know» that a wavelength of 490 Nanometre is called «green», but 480 Nanometre is «blue»? How does the system «know» that leaves are green? The simple answer: it does not «know» any of those things, these assertions are put into the system in one way or the other by humans. Modern machine learning systems use the «wisdom of the crowd» to obtain this information. But ultimately, the software calculates numbers and applies statistical methods.

To be clear: The results are impressive—that is, they are impressive as long as you do not count the false positive or false negative decisions. If the given problem is deciding whether a brushstroke is green or blue, there is no deeper moral dimension to it. If the given problem is to decide whether you will get a job the moral dimension is clearly visible. For the computer system both problems are essentially the same. It is us who have to decide when and in what context to use a data-driven decision-making system—or to decide not to use any assistive systems at all.

2.5 Algorithms, demystified

MENACE was the name of a machine built by Donald Michie in the 1960s that could play Noughts and Crosses (also known as Tic-Tac-Toe, Three in a Line, or Tateti) against a human player (cf. Michie 1961). The Machine Educable Noughts And Crosses Engine was a machine learning system but with a twist: The machine was made of matchboxes filled with coloured beads. The setup was quite impressive, no less than 304 boxes are needed, one box for each possible configuration. MENACE makes the first move by picking a matchbox labelled with an empty playing field, shaking it, and drawing a coloured bead. Of course, it needs an operator performing these tasks. Each colour represents one of the nine possible positions a X or an O can take on the playing field. In the course of the

first games the machine will likely lose because there is not strategy whatsoever involved, because the beads are drawn randomly. Enter the machine learning part: If MENACE loses, the operator will remove all the drawn beads that lead to defeat. If MENACE wins, the operator will add three beads of the drawn beads in each picked box. That means that the chance of losing again will be reduced while on the other side good moves are rewarded considerably. Trained long enough, MENACE will «learn» a winning strategy (by improving the chances of good moves) and therefore will «play» pretty well.

The interesting part is that no human player would ascribe any intention to a pile of boxes in contrast to machine learning systems implemented with software on a computer hardware. And even when there is no machine learning or «artificial intelligence» involved, a critical observer of the information society is still puzzled about the «enormously exaggerated attributions an even well-educated audience is capable of making, even strives to make, to a technology it does not understand.» (Weizenbaum 1976, p.7)

MENACE is a powerful didactic tool to demystify the process of machine learning. Yes, it is impressive, that a system can «learn» to play a certain game without knowing the rules (and not knowing whether it has won or lost because there is no entity capable of acquiring knowledge but just a bunch of boxes). But, no, this system won't be able to, say, play Chess. To be fair, you could theoretically construct a chess playing matchbox system, but for that game between 10^{43} and 10^{47} matchboxes are needed—needing more space than our solar system provides so it would be highly impractical. That is why we use computer systems; a Bit (binary digit) doesn't take up much space (but it *does* take up space). MENACE is not a metaphor for a machine learning system, it is a machine learning system. Of course, there are much more sophisticated machine learning systems out there, but all have one thing in common: They were designed and created by human beings with a purpose and applied in a specific domain. Outside that domain or used for any other purpose, these systems are useless (in non-critical contexts) or harmful (if used for sovereign tasks). So, MENACE is a good way of thinking about the use of decision-making systems: Do you want a pile of matchboxes «decide» what to wear today? Probably, I mean, why not. But

do you want a pile of matchboxes «decide» whether you will give a job to someone? Definitively not.

3. Mind the Data Gap

In this last section I would like to focus on the emancipatory aspects of data, computation, and algorithms. It is tempting for articles that critically examine technical development to speak of cultural decay and to mention only technological abstinence as the only way towards a sustainable world. But this would only reinforce the impression that there is no alternative to unlimited technical growth and would once again give technology an agency. Technology does not act; people act. They act on very different levels: As individuals, as a group or as an organisation. But what is the ultimate goal of all actions combined? This is one of the big questions posed long time ago and this is precisely the question that humanity has been trying to answer ever since it was able to ask questions. The most promising attempt at a response by contemporary humanity is called *conviviality* (cf. Illich 1973). In this sense, data should be usable by all people as a tool for working and living in the Information Age beyond the paradigm of producing something. Moving from productivity to *conviviality* means putting an ethical value above economical or technical value. Knowledge plays a central role in the process of negotiating a path towards a *convivial* society. As indicated above, there is a close link between knowledge and technology, and one key to this link is data. Data can make facts visible or conceal actual circumstances. Data-literacy is one of the pillars of a yet to be determined basic education of the digital society. The data scientist Hans Rosling was committed to providing data, facts and statistics to humanity for a mutual understanding of all cultures of the world. In his magnificent lectures and podcasts, he also exposes an educated audience for not including current figures and data in their assessments of the situation of the nations of the world. On stage, he uses a special visualisation developed by his Gapminder Foundation that shows how important the visualisation is when it comes to data, especially considering trends. And of course, the availability and accessibility of data and computational power is important if we strive for an inclusive algorithmic society.

3.1 Learning data literacy by making data

There are many fields where there are not enough data. The Gender Data Gap described by Caroline Criado Perez (2019) for example shows that some data are simply not collected. Data sets were and are still created with a male perspective in mind. This can be generalised: Data sets are created to benefit groups that already have economic power over other groups for the simple reason that good data are expensive. You have to invest in data analysis and processing and therefore you expect a return of interest: data as a means of payment.

There are other types of incentives other than money, for example political and informational power. The main incentive of Luftdaten.info, to pick a well-known civic tech example from my country, was to collect environmental data to shape the public discourse about fine dust in Stuttgart (Germany). Before this project started, there were no data available to the public regarding this important environmental issue. That is no coincidence: The car manufacturers in Stuttgart have a combined revenue of 200 billion EUR, if you add to this the automotive parts manufacturers with a total revenue of over 100 billion it should be clear that this power affects the public discourse when it comes to fine dust pollution. In 2015, Luftdaten.info (now called sensor.community) was founded as a civic tech project to measure particulate matter and fine dust. The team created a how-to for a low-cost sensor kit that could connect to an online service to collect and share the data in an open and accessible way. The initiative received attention regionally, nationally, and internationally when in 2015 the European Commission threatened to take the case about air pollution in Stuttgart to the European Court of Justice (ECJ) if Stuttgart's citizens are not quickly and effectively protected from excessive particulate matter concentrations. Since Luftdaten.info is a community project propelled by volunteers with a core team of ten people, the costs are extremely low (.000002 billion EUR, mainly for infrastructure). The lack of public data infrastructure was compensated by creating an open data infrastructure that is used by the public. For example, German Universities hold data science challenges («data hackathons») on a regularly basis using Luftdaten's data sets. The low barrier technology needed for a sensor kit means also that this kit is used in schools and educational projects.

Bridging the data gap with self-made gadgets and tools creates environmental awareness and political participation in a very playful way: Learning by making. Now, in the second age of the Turing Galaxy, the *technê* of data is more important than ever to maintain informational sovereignty, what was true 30 years ago has now become even more important:

Information technology does not bring us a land of milk and honey either, not even an informational one, in which the cooked information, as it corresponds to our appetite, flies into our eyes and ears. Not only does it mean much more trouble with information, because the information level is constantly rising, but information technology also demands constant superiority over its systems.[...] The tremendous expansion of world knowledge (“Realwissen”) has an obvious dark side: it is impossible to possess it as an individual and it is becoming increasingly difficult to orientate oneself in it beyond the immediate world of experience. [...] And the computer holds the danger that the broad masses, who set the tone in democracy, will only stick to what can be called up on the screen. (Zemanek 1991, pp.275-277, own translation)

3.2 Final remarks

Technology frees the forward-thinking human being from the world of necessities and constraints in a wonderful way, as the choir of the Theban Ancients sings at the beginning of the second act of Sophocles’ *Antigonae*. The technician ploughs the field, travels with horse and ship, catches birds, inhabits the highest peaks, invented (sic!) language. In short, he (and in ancient times the male pronoun is used very unambiguously) has more of *technê* than he can hope for, but unfortunately not enough of wisdom. Data is now part of our— and here I mean all people regardless of gender—common language, a *technê* that could help to speak the truth. But data as a modern *rhetorike techne* can both distract from the truth or be its revealing tool. Any participant in the data-driven discourse always pursues political intentions (especially the self-acclaimed *parrhesiastes*); data cannot be anything other than political. Data literacy is political education.

Bibliography

Bourne, C. P. (1963). *Methods of Information Handling. Information Sciences Series*. New York: Wiley.

Capurro, R. (2019). Enculturating Algorithms. *NanoEthics*. 13, 2, 131–37. doi: <https://doi.org/10.1007/s11569-019-00340-9>

- Coy, W (1994). Die Turing-Galaxis. *Report of the computer science faculty of the University of Bremen*, 3/94, 7–13. Bremen: University of Bremen.
- Perez, C. C. (2019). *Invisible Women: Exposing Data Bias in a World Designed for Men*. New York: Random House.
- Illich, I. (1973). *Tools for Conviviality*. New York: Harper&Row.
- Kiermeier, F., Renner, E. (1960). Einsatz der Randlochkarte in der Arbeitskartei (ein Beispiel aus der Milchwissenschaft). *Zeitschrift für Lebensmittel-Untersuchung und Forschung*, 113, 4, 316–322. doi: <https://doi.org/10.1007/BF01353946>
- Latour, B. (1990). Technology is society made durable. *The Sociological Review*, 38, 1_suppl, 103–131. doi: <https://doi.org/10.1111/j.1467-954X.1990.tb03350.x>
- Lovelace, A. (1842). Notes, In: Babbage, H. P. (Ed.). *Sketch of the Analytical Engine by L. F. Menabrea*. p. 21–50. New York: Cambridge University Press.
- Michie, D. (1961). Trial and Error. In: Barnett, S. A., McLaren, A. (Eds). *Science Survey*, Part 2, 129–145. Penguin Books Ltd.
- Orwat, C. (2020). *Risks of Discrimination through the Use of Algorithms*. German Federal Anti-Discrimination Agency. Baden-Baden: Nomos. Available at: https://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/EN/publikationen/Studie_en_Diskriminierungsrisiken_durch_Verwendung_von_Algorithmen.pdf?__blob=publicationFile&v=2
- Parry, R. (2008). Episteme and techne. In: Zalta, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. (First published Fri Apr 11, 2003; substantive revision Sun Oct 28, 2007). Available at: <http://plato.stanford.edu/archives/fall2008/entries/episteme-techne/>
- Pohle, J. (2016). Personal data not found: Personenbezogene Entscheidungen als überfällige Neuausrichtung im Datenschutz. *Datenschutz Nachrichten*, 39, 14–19.
- Turing, A. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society*, 42, 230–265.
- Ullrich, S. (2019a). Algorithmen, Daten und Ethik. In: Bendel, O. (Ed). *Handbuch Maschinenethik*. Wiesbaden: Springer VS. p. 119-144
- Ullrich, S. (2019b). *Boulevard Digital: Öffentliche Meinungsbildung Der Hypervernetzten Gesellschaft*. Wiesbaden: Springer.

UNESCO Institute for Statistics (2020). *Pupil-teacher ratio in primary education* (headcount basis). (14 September 2020)

Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgement to Calculation*. New York: W. H. Freeman.

Zemanek, H. (1991). *Das geistige Umfeld der Informationstechnik*, Berlin: Springer.

Author's notes

Acknowledgements: I would like to thank Rafael Capurro for the opportunity to publish my thoughts here within this group of wonderful experts. I would also like to thank my supervisor and mentor Wolfgang Coy, who brought me to the ethics of computer science. And last but not least, I would like to thank my family for their support, especially towards the end of the deadline for this article. Thanks Andrea, Carina and Leonard!

Disclaimer: This article uses Ullrich (2019a) as a main source. I have kept some of my favourite formulations and arguments.

Financing note: The Weizenbaum Institute for the Networked Society is fully funded by the Federal Ministry of Education and Research of Germany under grant no. 16DIII11.

Author contribution

The entirety of this manuscript was prepared by Stefan Ullrich.

Editor's notes

The editor responsible for the publication of this article was Rafael Capurro.

Style editing and linguistic revision to the wording in this text has been performed by Prof. Adj. Hugo E. Valanzano (State University, Uruguay).

Nilzete Ferreira Gomes (Universidade Federal Rural da Amazõnia (UFRA), Pará, Brazil), was in charge of translating from Portuguese to Spanish.