TOC non Technical

Control Lost

Tech going loose

Contents

etc. The new business models. Where have IBM and Microsoft gone? Platforms. Frontend, Databases, Frameworks. The role of Facebook, Google, Twitter, Uber

6- US Losing Control: Innovation: Where? US: Silicon Valley, Investments (> 1B$ in California, otherwise woman of color owned business), and Government-funded Research. Internationalization, Outsourcing and Immigration: India, Ukraine, Rumania, China. Europe and Regulations. US Government Losing Control: How to make it work from inside tech. Passive Government. Congress and Technology. Government ought to do something. NIT.

7- Remedy: NIT

Feeling of inevitability, immutability, immunity Rules of SEO similar and multiple checking comparable to tax return constraints. Notion of net states vs nation states (Alexis Wichowski) Feeling of powerlessness.

TOC technical

Data Networks

Targeted Audience:

- Professionals, Data management
- People dealing with big data, need to improve quality
- Government web sites
- Healthcare data professionals
- Legal data professionals
- Investigation: Law firms, law enforcement, investigative journalism
- Librarians
- Archivists

Topic Maps

History of Tax Map

Problems

- Hyperlink limitations
- Taxonomy rigidity
- Machine-processing vs. human processing

Solutions

- History
- Single source of truth
- Overlay
- Multipoint mapping vs. n-to-n mapping.
- Tax Map Requirements:
    - No change to existing workflow
    - Automation of dynamic publishing

Perspectors

Problems

- Auditing
- Accounting / Accountable Information

Solutions

- Processes = Semantic Links

The Networker Documentation

Problem

- Need for human curation
- Black-boxed algorithms: AI / ML

Solution

Synopsis

Data gone awry

The Technology Ideology

Defining technology ideology. Should be in one page. Scattered all over.
The fix. What is the fix? Is the fix applying a different technology?
Is this fixable? If there is a solution, it is likely to follow these lines. That is a good way to segway.
Junk food. What is the solution? That is what I could address.

Michel Biezunski, New York, February 2022.

Introduction. It does not have to be like this.

Technology plays an essential role in society, and the industry has concentrated into the hands of a small number of giant corporations which have accumulated an unprecedented amount of wealth and power. Technology has improved many aspects of our lives, but has also created new challenges that impact not only our well-being and mental health, but also the society in general. Data are also used for nefarious purposes. Cyberwars have become a real threat and attacks create serious risks of disruption of economic activities. Political propaganda is channeled using the same platforms than social networks. The ability to target recipients for specific information reinforces polarization and isolation.

Moreover, the enormous amount of data being extracted, collected, sold, and distributed creates a sentiment of being overwhelmed. We have more data to consume that we have been prepared for, and it diminishes our ability to take responsibility for our actions and our decisions. This affects not only our personal lives, but also our professional lives. People working in companies tell their customers reporting issues that even if they understand their problems, there is nothing they can do about it, "because it's the technology". When this attitude gets generalized, it affects the fabric of the society in general, and more specifically the core of democratic values, based on the ability for individuals to express their voice and play a role in how a country is run.

People who care feel powerless when confronted by problems they acknowledge but don't know how to fix, as if they were screaming in the void [^dont-look-up]. They suffer from addiction, they take it personally, and try to act on it at their individual level at the cost of a considerable effort. The span of the crisis in the society is comparable with obesity: in both cases, there is an elephant in the room. The food industry has its share of responsibility in the obesity crisis, as companies are producing junk food at high profit margins, regardless of the effects on the health of their consumers. Similarly, the social media platforms and other big tech companies are creating highly addictive products at very competitive prices -- most of the time, for free -- regardless of the consequences for the well-being, dignity, and mental health of their "users", and on the collective effect on the society.

[^dont-look-up]: The success of the movie *Don't Look Up* reflects the shared feeling of not being heard.

The age of surveillance capitalism[^zuboff] is made possible by the fact that data collectors stay free from regulatory constraints that would make it much more difficult for them to continue amassing these data. The European Union has been proactively advancing legislation to that effect, but it only has a marginal role on the politics of technology companies, which are mostly based in the United States of America, while the electronic product manufacturers are located in Asia. The business model based on collecting user data is being opposed by other tech giants, which consider that in the long term they are better off being respectful of their customers' interest rather than alienating them.[^apple-privacy] Apple for example seems to be determined to protect its customers' privacy by preventing Meta from collecting as much data. Trouble lies ahead for companies which already had to change their name to a non-meaningful name, i.e. a name that is used in so many different contexts that it has lost its dent. Although "google" is now a verb, the company considered it a bad name and changed it into "Alphabet", and Facebook went for "Meta". [^alphabet-meta]

[^alphabet-meta]: *Alphabet* can be used for anything (at least in the Western civilization) and *meta*, similarly is a prefix that can be attached to diverse concepts. It is interesting to note that neither Apple, Microsoft, nor Amazon, who are less dependent on surveillance, didn't feel the need to change their name. This may be a sign that the page is being turned.

[^apple-privacy]: Anupam Chugh, "Apple is Killing a Billion-Dollar Ad Industry With One Popup", *Medium*, July 7, 2020. https://medium.com/macoclock/apple-is-killing-a-billion-dollar-ad-industry-with-one-popup-2f83d182837f and Sylvain Saurel, "The Clash of the Titans To Come in the Metaverse: Apple vs. Facebook", *Medium*, Jan. 17, 2022. https://medium.datadriveninvestor.com/the-clash-of-the-titans-to-come-in-the-metaverse-apple-vs-facebook-1d48bbd9b36a, and Justin Kruger, "Is Apple planning another Big 'F-U' to Meta's Metaverse, by fully supporting WebXR Open Standards, *Medium*, Feb. 8, 2022. https://medium.com/@jdavid-net/is-apple-planning-another-big-f-u-to-meta-by-fully-supporting-webxr-open-standards-9974f3da78ae

The Big Tech giants are trying to convince people that their way is the only way. To that effect, they are entertaining a system of beliefs that rejects its critics as

Luddites. "Privacy is dead. Get over it", is the famous sentence by Mark Zuckerberg that sums it all. Most people are consequently convinced that losing privacy is an inevitable consequence of using technology. It is not. It is possible to search the Internet without being tracked, as alternative search engines to Google, such as Duck-Duck-Go, prove. Similarly, nothing prevents to design social networks that do not store users' data. The only reason why consumers' data is sold to the advertisers' industries is to increase profit from technology companies, or to reinforce the levers of control of authoritarian regimes on their citizens. It is not something they have to do in order to make their technology work. They only do it because they don't encounter obstacles. They claim that no technology would be possible without tracking. This is as preposterous as pretending that the Post office has to open letters sent by mail in order to send it. They don't. They use stamps, and as they are considered to be working for the public good in general, they have been sending mail regardless of whether they were making a huge profit or not.

The technology ideology is this system of beliefs that big tech giants want us to believe in, in order to continue to maximize their profit margin. They exploit the fact that technology is inherently complex, it is made of many layers. At the beginning of the Internet era, it was possible for one person to get a good knowledge of the diverse layers involved. Today, it has become almost impossible, and the domain of technology is split into many sub-specialties, and it has become practically impossible for one person to extend their expertise to cover all the layers.

Data collection has increased so much that it has disrupted the traditional ways to handle data: the classic data handling techniques, consisting in organizing, classifying, connecting individual data points, search for meaning, etc., are getting discarded and replaced by algorithmic techniques on which most people, including many programmers, have no power. Artificial intelligence and machine-learning have become buzzwords that fit well with the universe of Big Data.

This book is about data. The shared vision about data is that there is so much of it that we need to rely on algorithms to collect and navigate the data.

- Why is there so much data?
- Who needs that amount of data?

- Small data is beautiful. Big Data is not all there is.
- Is data only valuable as a statistics? Aren't there situations in which we need data to be precise, accessible, and controllable?
- Who controls the accuracy of the data?
- What about data being findable? What methods? Shouldn't it be open? Should we know what rules are being applied? Who makes the rules? What if our rights are violated (for example, we're not found.) It is possible to imagine situations where everybody would be able to publish (first amendment rights), but nothing would be found.
- Are political campaigns the same as advertising?
- Is small business advertising on Facebook the only window available?
- What about laws governing publishing? Libel laws, for example. Violations of privacy rights?
- What about universal agreements? (Upper ontologies, Linked Data, Semantic Web, XML schemas, Universal databases?)
- Is the value in the nodes or in the links?

Solution:

- Fight against the ideology of inevitability. If you're not with us, you're a Luddite. Examples: Fran Leibowitz, Richard Stallman (not exactly a Luddite).
-

A problem well identified is easier to address. The main problem we are facing is the fact that the technology industry has created an ideology, that they want everybody to believe in, that basically asserts that things are what they are, because that's the inevitable way forward, that technology can't be avoided, and that any attempts to tame it are caused by Luddites who want to go backward because they do not understand it, as would those who advocate to returning to candles instead of using electricity.

The food industry is slowly changing course now that many people have realized that it is possible to purchase food that is also good for our health. Even the big food corporations understand that there is life after junk food, and they may even make profit in this changing context. But the technology companies are so entrenched

with their current business models that they want all of us to believe that their way is the only way. They have created an ideology which is a system of beliefs they want everybody to share, for example claiming that "privacy is dead, get over it". Any vision that counters it, they claim, can only be uttered by Luddites who are looking to go backward, and want to return to candles instead of using electricity. However, there is no direct relationship between searching on the Internet and being tracked. The only reason this is related is because it justifies the business model of big tech corporations, which resell the data provided, unknowingly, by their users. The ideology consists in convincing everybody that one can not go without the other.

The fact that we are led to accept that essential aspects of our lives -- our integrity, our privacy, our intimacy, our dignity, our connections -- are not any more under our control has numerous consequences, especially on the morale. It has changed our behaviors, and also our brains. The generations born in a world permeated by social media have a different perception of what it means to live together, to relate to common values.

This affects not only individual owners of smartphones who spend most of their time on social media platforms, but also professionals who are dealing with data. There is so much data available that overwhelmed by the amount of data they have to deal with, and, are yielding their responsibilities to algorithms they don't understand, under the shared common understanding that the human brains are not able to process efficiently as much data as machines.

Big Tech companies are enjoying a level of wealth and power that is unprecedented. Their influence makes them equal to -- or even in control of -- national governments, as their decisions influence the life of every individual and every institution in every country of the world. Their goal is to continue to expand. In order to achieve this goal, they need to make sure that their domination remains unchallenged. They have created an ideology, that I will call the "tech ideology", that is a system of beliefs compatible with their business models. They deploy tremendous efforts to make it pervasively accepted by most influential decision makers, who in turn propagate it until everybody is convinced that this the only way to go.

This book aims at shedding a critical light on the tech ideology in the domain of data.

The freely available social media platforms providing ways for anybody to connect with anybody else, to publish anything that is immediately accessible worldwide, are powerful and addictive tools, that no one could have predicted or dreamed of in the 20th century. Similarly, the immediate access to any information available on the World Wide Web through search engines provides tremendous advantages to those who use it.

There is no free lunch, as the saying goes. Those who use these fantastic tools for free are, often unbeknownst to them, giving away a lot of personal information that is collected for commercial or mere surveillance purposes. If they live in a country where these data are only collected for advertisement purposes, they received recommendations for products that have been calculated to be of interest to them. If they live in an authoritarian regime, the data collected is used to track any sign of disloyalty to the rulers, and they may suffer consequences. If the data is collected by people with nefarious purposes, they can be attacked, bullied, harassed because of information that they have posted online without thinking there could be such consequences.

Shoshana Zuboff[^zuboff] has written a masterpiece where she shows very clearly why the new business model of big tech has been so successful. She also shows that it was developed somewhat accidentally. It was not a given that Google should track its users. At the very beginning, they were not doing it, they didn't even think of it. Their search technology has nothing to do with it. It is only because they were wondering how to monetize their technology that they ended up thinking that the information they were able to get by basically collecting the trash left inside the computers was extremely valuable. Not only they would provide their users access to universal knowledge, they could also provide their customers -- the advertising industry -- universal knowledge of their users. Facebook replicated that model by providing ways for people to connect with each other, therefore offering a free publishing platform, and at the same time providing them with an intimate knowledge of every of their users, which constitute a significant portion of the world population.

The more data they can get their grip on, the better for them.

Alexis Wichowski[^wichowski] calls those big tech companies "net states" because they are playing a role as makers and destroyers of democracies, as well as supporting authoritarian regimes, as the impact of their decisions weigh often more than applying a law. In the United States the ability or not to gain access to a social media platform plays a significant role in the political landscape, especially when the content published is considered controversial. Authoritarian regimes have well understood the benefits they can gain by reining in technology to their interests. When Facebook came up with the "Like" button, it was perceived by their users as a nice way to communicate with their families, friends, colleagues or neighbors. It was used by Facebook to acquire knowledge that can be used to improve the quality of the targeted audience they were selling to the advertisers. The exact same mechanism is used by China to ensure that their subjects are expressing loyalty to the regime, and to punish those who don't.[^wichowski2]

Technology is also used in the public interest. The pioneers of the Web envisioned a web where communications between human beings would be greatly facilitated, access to knowledge and education would be easier, frontiers would be abolished. In the first years of the Web, there was a hope that the world would become a better place with new, open technology. Then the Open Source community has continued to develop powerful tools that are massively used and have made a big difference in the world.

In his book, entitled *Runaway Technology: Can Law Keep Up*[^fairfield], Joshua Fairfield demonstrates the fallacy of the claim that law can´t keep up with the rapid pace of technology, as he explains that law itself is a technology. The web has born by implementing a technique, called hypertext, which mimicked the century old legal tradition of referencing, used for statutes and cases, on the pre-existing Internet digital network. Joshua Fairfield unravels the ideology that is created by the major technology players in order to reinforce their power, shut down their critics, and justify their overarching influence.

Whether the initiative belongs to big corporations or to the authorities, the common outcome is that the people -- relegated to being called the "users" -- have no voice.

The opponents are labeled as Luddites and are seen with contempt as small fry of nostalgics of a past era, soon to be forgotten: "Privacy is dead, get over it" is one essential component of this ideology. It is based on the idea that technology inevitably leads to changes that no one can overcome. However -- and this is why it is just an ideology -- it is perfectly conceivable to imagine a social media platform that does not track its users, or an email platform that guarantees that the content of the messages will not be open for reselling information contained in them, as is the case for mail handled by the Post office. Nevertheless, it is more difficult to figure out how to make astronomical profits out of a service that benefits the public while *not* exploiting private data for commercial purposes. The decision by Apple to let users opt out of being tracked has proven disastrous for FaceBook.[^apple-facebook]

[^apple-facebook]: Meghan Bobrowsky, "Facebook Feels $10 Billion Sting From Apple’s Privacy Push", *Wall Street Journal*, February 3, 2022.

In a far future, our descendants who still live in democratic societies may communicate solely through electronic channels without fearing for their individual freedom and security. But we still have a long way to go to get there. We are too entrenched in the old ways of thinking to be able to tame the newcomers in the tech world who have seized any available data. In order to advance, the first step, obviously, is to understand what is happening.

There are numerous initiatives for advancing the rights of citizens in the digital era. Local governments, for example cities, are actively promoting more inclusive policies. The concept of public interest-oriented technology makes its way. The open source community, which originates from the glorious period of the nascent Internet and early web, is still around and active.

In this book, I will focus on data and show how the technology ideology handles it. I will specifically show that the concept known as *big data* aggravates the publicly held sentiment that things have gone out of control in the world of technology. Starting from the innocuous observation that the mere amount of data available is growing at a ferocious speed -- it is no wonder if we feel overwhelmed and powerless in the face of it. As technology gets more and more complex, the general

feeling that handling data is out of reach for the ordinary mortal is growing. Therefore, outsourcing and giving a blank check to the technology company seems to be the natural way to go, and is the preferred solution, for many of the people who are in the business of dealing with data.

[^wichowski]: Alexis Wichowski, *The Information Trade — How Big Tech Conquers Countries, Challenges Our Rights, and Transforms Our World*, Harper Collins, 2020.

[^zuboff]: Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Public Affairs, 2019. The author shows that the big tech corporations are working hard to keep their power unchecked, regardless of the consequences for the rest of the society. Their business model, based on extracting and reselling data provided by their customers, has enabled them to acquire a level of power and influence traditionally only reachable by nation states.

[^fairfield]: Joshua Fairfied, _Runaway Technology — Can Law Keep Up? _, Cambridge University Press, 2021.

[^wichowski2]: A. Wichowski, *Ibid*, Chapter 5: "A Great Wall of Watchers", p. 131-152.

Part 1. The Problem. The Technology Ideology

Two kinds of big data

There are two kinds of big data: first, there is data which size has increased gradually, as more data got integrated into the mix. For example, product catalogs of big online sellers, integrated encyclopedias such as Wikipedia, legal and administrative information made available by governments worldwide, geographical maps. What these data sets have in common is that, despite their ever-increasing

size, they remain manageable -- and they are being actively managed. At an individual level, the number of personal data, such as photographs, has increased at the point that they seem overwhelming. However, there are ways to deal with it, providing we invest the time and effort.

Second, there is this trove of data obtained by tracking usage data. The sheer amount of data is much more extensive than in the first case. It keeps growing exponentially at every minute. What is recorded is, for each individual, their whereabouts, their relationships with one another, their expenses, their search history, their clicks, their professional careers, their travels, their expenses, their daily schedules and calendars, their political beliefs and religious affiliations, the content of their conversations and messages, their heart rates and other medical conditions.

These two categories of big data are quite different. The organic integration of knowledge imposes requirements, such as maintaining a high quality, accuracy, and reliability of the data. It also requires that the origin of the sources be tracked, making the content accountable. On Wikipedia, it is essential, when reading some piece of information that looks weird, to be able to retrace the history of the page, and engage in a discussion with the author of those lines. Amazon and the companies engaged in e-commerce are committing to a precise product description, a certain price, and shipping times. If either one of this information is not accurate, the customer is likely to cancel the sale and may not even return to the company to buy products. Therefore, the maintenance of the accuracy of data still remains essential at the granularity level of each data point.

Data which is acquired by tracking is of a different nature. First and foremost, it is not transparent, as the people being tracked may not even know that their information is recorded. They have probably accepted to decline any claim to the contrary, in the fine print of the multiple page long user agreements that they click on in order to accept services. Data doesn't need to be precise nor interpretable at an individual level. Because it has value mostly when aggregated with similar data at a much bigger scale. The algorithms used for face recognition wouldn't make sense if trying to be understood by one person who would like to see their personal data. Face recognition software have a hard time identifying people wearing masks.

Why is big data so big?

The more data there is to be analyzed, the more accurate the analysis is. Therefore, companies which are in the business of selling data need not to have data where every data point is 100% accurate, but the need to get as much data as possible, creating the need for an ever growing amount of data. User tracking is possible as long as there is no obstacle. Shoshana Zuboff talks about the deals that the US governement made after 9/11 with Google, with Google committing to provide as much data as the government needed, with the condition that government would remove all regulatory attempts aimed at curving Google's ability to track data. The Obama campaign in 2012 renewed that deal. The visible abuses caused by Big Tech are being sidetracked as the classic monopoly issue, and solutions that are designed are made with the intent to break monopolies. As the issue of tracking is somewhat unrelated to the monopolistic behavior, this is purely a distraction: Google being not the default search engine any more would not be prevented to track online users the way they please, and would still retain their ability to resell users' information to the advertisers.

If instead regulatory provisions would be created that would prevent data to be acquired without the explicit consent of the persons to which this data belongs, that would severely limit the amount of data generated, and big data would therefore become smaller.

In other words, big data can be seen, in a historic perspective, as an abnormal swelling of the data amount due to the laissez-faire of regulatory institutions. It is not because it is technically possible to acquire data that data should be actually acquired. Why aren't we recording all atoms in a human body? Probably because the technology doesn't allow it yet, but if the current trend continues, we may eventually find reasons to do it.[^atoms] However, on a more somber side, the consequence of this universal tracking is that there seems to be no safe place to escape from being tracked everywhere. As long as the only reason for being tracked is to receive recommendations on what to buy, we can comfortably ignore it. But once authoritarian regimes use the exact same techniques to monitor every citizen's activities and to exert diverse kinds of punitive pressures on any dissident,

the incentive for making these practices accountable becomes much higher. It becomes an issue of essential freedom, sometimes of life and death. Therefore, those of us who neglect to act to protect themselves from being tracked do a grave disservice to those who are prosecuted for who they are, what they do or what they think. It is not as innocuous as it seems.[^hong-kong]

[^atoms]: Perhaps medical breakthroughs could come out of this activity. There are approximately $7x10^{27}$ (7,000,000,000,000,000,000,000,000,000) atoms in an average human body (male, 70kg), according to Anne Marie Helmenstine, "How Many Atoms There Are in the Human Body", April 5, 2019, ThoughtCo https://www.thoughtco.com/how-many-atoms-are-in-human-body-603872.

[^hong-kong]: Protesters against the takeover of democratic rules in Hong-Kong using sophisticated technologies, including facial recognition, have been unvoluntary pioneers in showing what technology can lead to.

There is an indirect, more insidious, consequence of the generalization of the usage of big data: it is the fact that the techniques used to handle data at a large scale are generalized to the handling of data at smaller scale, but still described as large. J. Fairfield explains that data science introduces a paradigm shift, in the sense that, instead of being a science that tests theories against data produced in experiments, it is a science that purely uses the data with no pre-existing theory. Data science consists in pure correlation between data sets that have been accumulated without any particular purpose in mind -- at least, explicitly.

For example -- and this is an imaginary, made up example -- we could establish a correlation between wearing black pants and buying a given brand of raspberry. Such a result needs not be supported by a rational explanation (why is it so?) but it nevertheless could be used by companies which sell pants and companies selling raspberries.

The technology ideology aims at maintaining the status quo, i.e. preventing any regulation to create obstacles to the accumulation of data through universal tracking. The big data handling techniques, based on algorithms using what has become known as artificial intelligence and machine-learning (or deep learning), have two simultaneous, and complementary features: they are used when there is

too much data to be handled by the classic methods based on human expertise, classification, taxonomy where things are categorized and grouped according to common features. They are based on this premise that artificial intelligence far exceeds the capabilities of human brains. By shifting data processing to machine processing, humans are asked to commit to a convenant of blind trust to the power of technology, renouncing to any critical view on the matter. Things are what they are, and we can't say anything about it. But things are getting worse. Instead of confining to the boundaries of "really big data", this vision has invaded also the domains that were traditionally less big in size, where humans were in control of organizing their data. Now, they are told, they have to switch to automated methods. It's going to be more efficient, more accurate, and they won't have nothing to do any more, as the machines will take care of everything. In reality, the essential reason for pushing for this radical change is to further reinforce the influence of big tech corporations, which are depicted as the only ones that "know what they are talking about", discarding any alternatives as remnants of a past era.

Once again, there are two main beneficiaries to that attitude: the big tech corporations which use their increased credibility to continue to oppose to any accountability in the way they acquire data, and authoritarian regimes, who are able to convince their citizens that these technologies based on AI and ML are going to leave them no place to hide. Resistance is futile, or so it seems.

Taking the opposite point of view, this book aims at showing that resistance is possible. In fact, the way we can build a better future in which technology can be used to improve people's lives, both at an individual level, and at the collective levels of societies can only pass through an initial phase of reckoning that resistance is needed against the mainstream belief that whatever is happening is inevitable.

Technology is in a middle ground. It is not neutral, because it can not be disconnected from the purpose for which it is being used. There is not one side that is for technology, and a side that is against technology. Rather, there are those who use technology to improve the wealth of big corporations, those who use technology to empower authoritarian regimes, and those who use technology to empower

citizens and strengthen democracy. Technology is not political per se, but it is used in different ways consistent with divergent projects for the society.

The ETL paradigm

Is artificial intelligence better than human intelligence?

The fact that the same term, intelligence, is used to designate the results of algorithmic processes is part of the technology ideology. There is nothing "intelligent" about a way a computer processes data. It is programmed to produce predefined results which varies according to the nature of the data that it is being fed with. Classic algorithms are producing new data by logically combining input data. AI/ML algorithms are not producing data. Computers are even able to automatically write algorithms, a procedure known as "machine learning". By ingesting more data, the results are more accurate.[^fairfield] The products of artificial intelligence processes is a set of consolidated data, presented in a way that is digestible by human brains. Even if computers take some decisions, at the end of the day, humans are still the recipients. They use their brainpower to exploit the data. Therefore, excluding human intelligence from any computer-based process is absurd.

Computers are able to process lots of data more efficiently and more accurately than humans, but there are many things that computers are not able to do: exert critical thinking, think out of the box, invent new ways of doing things, accept ambiguities, compare multiple perspectives, detect blunt errors, detect whether something is missing, wonder whether something is ethically desirable or not, forecast potential side effects, weigh positive versus negative consequences of decisions. Practically all of these characteristics typical of intelligent human decision making are simply absent from Artificial intelligence, as if they would not count.

In their absence, what remains is just one way of doing things. If it is flawed from the beginning, it is flawed at the end, and any bias introduced in the process is multiplied and propagated like metastasis. Only coercion towards the unique way of doing things can succeed in getting every one's agreement. People who work within such a closed environment are enslaved, as they are not able to weigh in to modify the system to fit what they think should be done. The call centers assistors start to behave like robots. They keep repeating over and over the response that they are missioned to give. If the consumer insists this is not what they expect, the call centers employees answer -- sometimes -- that they understand the problem, but there is nothing they can do to fix it, because the "system" won't let them do it. Micromanagement, when it's performed by humans, can be counter-productive. When a computer records every key stroke, then the company gets statistical analysis automatically built that can help them decide, based on a set of numbers, which employee is less productive, and it helps them downsize quickly. In a normal situation, people whose work has been transformed into an endless activity of filling forms on screens, feel much less committed, and are losing their ability to even think whether what they are doing makes any sense or not. Even if some of these techniques bring a productivity boost on the short term, it can have devastating consequences on the long term, as the staff loses any feeling of commitment, any creativity and any hope that things are going to improve.

When a company decides to move the handling of their own data to AI and ML based techniques acting as black boxes, they lose any leverage they had over their own data and enter into a dependent relationship towards the company that provides the software tools that make the system work. But they are on a tight rope: if the quality and precision of the data is essential to what they deliver, they can not afford to ignore the details. For example, if they rely on proprietary search algorithms on which they have no control that do not return the items that their customers are likely to look for, they are cooked.

The main argument used by proponents of artificial intelligence is that Big data is too big to be processed directly by human beings. As more and more data is processed, machine learning algorithms are added to the mix. These are automatically generated algorithms taking into account the results of processing,

running them against the expected result, and refining the algorithms so that, next time they are used, they will be able to benefit from the "lessons learned" from the data they have already processed. Machine Learning, or Deep Learning technologies use advanced mathematical techniques, that very few people understand. The more they are used, the blacker the boxes become. The algorithms get conferred magical power.

Lost Control. Increased workers' exploitation

The obfuscation of the internal processes and the alleged magical power of artificial intelligence are an essential ingredient of the technology ideology, as it requires to engage in blind trust into technology and to reckon that humans are not equipped to confront those. Accountability is relegated to an obsolete requirement, at a time where forensic analysis was imaginable. Hence the sense of powerlessness and resignation that follows. And more power to Big Tech or to the authoritarian leaders, depending in which context those technologies are deployed.

Fixing problems becomes quite a challenge. When blunt and obvious errors are detected, everybody dodges responsibility, as nobody is really in charge. And this is where the user agreements come handy. Companies deny liability, if anything goes wrong, in the fine print. Users have been warned. These terms of use are a caricature of agreements, as their content can generally not be changed. The unfortunate consequence of this situation is that people who are actively using the technology have anything that resembles freedom of speech, as their voice doesn't count.

However, things are not that clear cut. Processes that appear to run smoothly through automated algorithms often employ an army of low-paying employees, "data curators", who are paid to prevent the companies that output these data to be sued for inappropriate content. They spend their time correcting the most egregious data points, including calls to violence, sexual exploitation of children, illegal trafficking, money laundering, hate speeches, calls to murder, libels, etc. Not only

these jobs are not paid well, but they take a toll on the moral of the workers, who are all day long confronted to disgusting images, videos, or text.[^sara-brown] [^steve-lohr][^steve-lohr2]

[^sara-brown]: >> [Not sure this is relevant] See Sara Brown, "10 big data blunders businesses should avoid", *MIT Management*, Sloan School, July 14, 2020. https://mitsloan.mit.edu/ideas-made-to-matter/10-big-data-blunders-businesses-should-avoid

[^steve-lohr]: Steve Lohr, "Algorithms Get a Human Hand in Steering Web", *The New York Times*, March 10, 2013. https://www.nytimes.com/2013/03/11/technology/computer-algorithms-rely-increasingly-on-human-helpers.html

[^steve-lohr2]: Steve Lohr, "If Algorithms Know All, How Much Should Humans Help", *The New York Times*, April 6, 2015. https://www.nytimes.com/2015/04/07/upshot/if-algorithms-know-all-how-much-should-humans-help.html

Data Analysis vs classic scientific paradigm. Are we done with science?

> > [Duplicate] Data science is tautological. It is perfectly possible to derive from data analysis that white males wearing blue denims are more likely to drink whole milk than teenagers of color in poor urban neighborhoods (this is an imaginary example). Why is that? Because the data only show something that was considered meaningful, for whatever reason, to start with. What is the purpose of that data? It can be sold to milk producing companies to help them target their customers where they are more likely to be, and save them money by not trying to penetrate markets in which their chances of succeeding is reduced. And the companies that accumulate this data can potentially make a significant profit by selling this data, showing numerical results that will help the executives in the milk company to reorient their

> sales forces. But at the end of the day, what have we learned? Practically nothing. Because the very fact that the milk consumption is part of the data set means that there have been some prejudices or a pre-existing market analysis, establishing for a fact that a market is more likely to be favorable than another.

> [Duplicate] As Joshua Fairfield mentions in his book Runaway Tech, there is a paradigm shift between a science that produces data and a science which starts from the data. Algorithms are of a different nature in both cases. In the first case, algorithms apply some logical reasoning based on explicit assertions to create the corresponding data. In the second case, algorithms simply aggregate existing data without any pre-existing hypotheses. Data aggregation consumes an enormous amount of data. For example, face recognition algorithms only work if there are millions of faces recorded, and a face is analyzed against those millions to be identified. The failed efforts of machine translation were based on the insufficient hypotheses about the structure of languages, hoping that grammatical parsing would allow to create a brand new translated version in another language. Using data-based algorithms, the translation of a sentence consists in trying to identify the same sentence, or fragments of it, already translated.

AI algorithms are an important piece of the tech ideology. They tend to reinforce the stereotypes that are illustrated by "real data", even if the data only show something that is the same as what was stated to start with. Therefore, if the data is meaningless to start with, it's as meaningless at the end, but it has received an "objective" blessing, as it is possible to assert that a statement is illustrated by data. It is a trick similar to the easy one, often used, consisting in inserting a mathematical formula totally unrelated to a question, in order to insist on the "scientificity" of the assertion.

Deep learning takes the issue one step further. Algorithms are created by other algorithms, and enable the data to be "improved", i.e. more precise. The technological ideology gets reinforced by producing data that can be more accurate

as data gets accumulated and the reference repository gets larger and larger. It also aims at widening the gap between the allegedly limited human brains compared to the ever more powerful computer power.

Moreover, J. Fairfield reveals how the tech ideology enables considering science just as another opinion by stating that data science is a paradigm shift away from science as practiced otherwise. In non-data science, data is produced by experiments that are there to either prove or discard a theory. Either the theory follows the fact, and it is confirmed, or it does not, and it is discarded. It is more complicated than that, because there can be several interpretations of the same data, that can lead to different results. But data science is different. There is no theory to account for, data is the only thing there is. The more data is available, the more the description of data is considered accurate and valuable. Theories have been replaced by dashboards. You drive your own conclusions. There is no factual basis that can either prove or disprove a theory. There are no theories, just opinions. The more powerful wins.

Part 2. The Fix

Small data is beautiful

There are many companies that sell data that they publish or distribute with the products they sell. Government agencies, international organizations publish data that have an authoritative value and are used as points of reference. Their commercial success or their credibility entirely depends on the reliability and accuracy of the data. They are often creating their own data, and are held liable for the content. For example, a government agency that makes information available to the public has no business spreading wrong information. If the agency is big, there is potentially a lot of information out there. Does it mean that this information should be treated the same way as big data? That is what is claimed by the proponents of the technology ideology. This gives a significant competitive advantage to the

companies that provide those technologies. The price to pay here is similar than for individuals. The government employees have to accept to lose control over the way their data is managed. The algorithms used with artificial intelligence and big data are almost always proprietary, and not accessible by customers, even big customers as a government agency. The management of information is being outsourced. Therefore the government employees are not considering themselves responsible for the quality of their own content, and even the government agency can decline liability. A similar situation occurs in call centers: when a customer needs something that the assistor can't provide, the usual answer is: I understand what you need, but this is beyond my control, the computer won't let me do it.location at any time creates a trove of data, that can be - but does not need to be - generated. It is only generated in the specific purpose of selling this data to the companies

The Knowledge Onion

Knowledge is built as layers upon layers. Typically, a scientific work is built by referencing other previous works. There can be several ways to quote the same work, and multiple perspectives are allowed.

The way information is described depends on one individual's perspective. As an example, the index of a book describes the main concepts contained in a book. The index is made either by an author, or by a professional indexer. If a book were to be indexed by two different persons, the resulting index would be slightly different. It would reflect personal choices, and it would be difficult in general to decide whether one is better than the other.

The hypertext system deployed across the Internet that played an essential role for the Web to emerge, in the early 1990s, was simply using the referential system of pointing to something else that is the basis of the legal system since centuries.

Tailor-made algorithms: Search Engine and Page Ranking

Google algorithm is tailored for the whole web. If search is limited to a single data repository, custom search can be built that corresponds to specific requirements.

Someone's metadata is someone else'data: No Schema needed

Information can be messy. There may be several ways to express the same thing, and one thing can be expressed with multiple names, in different languages or by using synonyms. The structure of data sets that are integrated can be different to start with and need some kind of "harmonization".

Database schemas, XML, types, classifications, ontologies, taxonomies

The human brain operates by associations between thoughts, that are different for each person, and depend on their previous experiences. The associations are not necessary predictable. For example, a woman wearing a green hat can be associated with a bridge in Rome, Italy, for reasons that make sense for one person, and for that person only. No one else would do this association.

Is universal agreement possible? Let alone desirable?

Human communication can be complicated. There may be a difference between what a person says and the way it is understood by someone else. Sometimes, the difference can be subtle, sometimes it is negligible, sometimes it is significant. There are not always good, reliable ways, to tell what a person really understood what another person meant.

Human communication also relies on what is not said. Sometimes, what is left unsaid is more important, and tells more than what is said. This can be expressed by other means, for example the body language, but even the mere absence of a subject that is communicated is like an elephant in the room. There are situations where there is an expectation that something is understood when it is not said. For example, a functioning democracy relies on the implicit assumption that everybody accepts the Constitution.

See illusion of esperanto.

Everything is interconnected

Value is in the links. Knowledge Graphs

Taking data into accounts

Accountability. Parallel with double entry bookkeeping. Money doesn't appear from nowhere, any monetary transaction starts in one account and goes to another account. In data land, it is possible to conceive that data is like money. No data comes from nowhere, it is being created, transmitted, or displays through a process that takes places between one information account to another information account. An information account can be a web page, or any computer representation of an idea. It is what is called "topic" in the topic maps architecture. A topic is a computer-representable proxy for a subject that can be related through any number of other topics in a network. The fact that a data point is created can be also described as a relation between an author and the subject they are creating. Revealing the processes (for example, creation) of data points opens the possibility to make data accountable.

See BlockChain.

Chapter 1.

Losing Control over Data

1. Overwhelmed by data

Our societies get many benefits from technology, but technology brings also negative side effects. This book is about finding where the dark side of technology resides, analyze the effects and discuss how to alleviate them.

Data is everywhere and rules the world. There is so much of it that automated processes to extract and process data has become essential. The expression "Artificial intelligence" is used to describe those processes, and the term serves also for his proponents to claim that machines are superior to humans. Consequently, the social status of "knowledge workers" is being downgraded. The irony is that the traditional progressive view is based on the belief will make human labor less painful, and the reality is that although this is happening, at the same time, work has becoming more dull, less creative, and individual initiatives are drastically limited because of the dependency to technology. This generates a feeling of uselessness.

Technology is complex, and technological products are very complicated constructions made to enhance the user experience. However, most often than not, users don't really have their say in the design of the applications. The perverse consequence of this is that the user must learn to behave as they were satisfied, whether that's true or not. Instead of being in the driver seat and express requirements that translate into a product that corresponds to their needs, users should lower their expectations to fit the actual limitations of the software they use. The term "user" says it all. The power is in the hands of the technology providers, not their users. This generates passivity.

Information technologies generates all kinds of data, including tracking recordings. There is nothing that can not be traced and recorded: activities, locations, clicks, visited links. Companies record their employees usages of computers. Wearable devices, such as smartphones, record their location in a very precise manner, as well as date and times, and the wireless-enabled objects (refrigerators, beds, watches) report all kinds of data, including health reports containing the kind of information that is usually available in medical tests such as blood tests. The feeling of being constantly monitored in all aspects of life generates a feeling of resignation and inevitability. It seems that the price to pay to use technology is accepting to give away anonymity and privacy. This generates initially a discomfort, that later becomes numbness or indifference. If we can't avoid it, we are better off accepting it and ignore it.

Going online is like walking in a street. When we go out, we expect to be recognized in the street by our neighbors or acquaintances. If we want to stroll incognito, we we have to be creative and use a specific outfit designed for that purpose: for example, a combination of a burka, face mask and sunglasses. But that is not convenient nor socially acceptable way to behave. We are not supposed to hide when we are in an open space. This parallel helps us accept the fact that if we go online, we should not expect any privacy, since this is an open space.

However, things are not that simple. Scores of online activities need to be protected. For example, financial transactions on a bank account should not be open to the public. But, as in the street we run the risk to be hit by pickpockets, online we encur the risk to be hit by thieves who steal our online credentials to penetrate into our private sphere. Changing often passwords, using two or three step identifications are band aids that create more burden to the user, go a long way to protect us, but are in no way considered an absolute protection. If we are the victims of online attacks, the companies providing the services are often not considered liable, although there are exceptions.

Terms of service provided by online providers do not leave room for negotiation. These terms are unilateral and they can only be either accepted or refused. Disagreeing with some of the aspects of terms of service is not an option. Consequently, there is no real incentive in reading the fine print. This creates a

sense of imbalance, and the recognition that corporations that provide services have the power to protect themselves, not us.

Credit scoring is a complex game invented by the financial institutions that rely on an average expected behavior. A good expected behavior does not consist in maintenance a healthy financial balance by having zero debt, but to manage debt in a way that both maximizes the profit for financial institutions, which charge interest rates, and show overall that the beneficiary of credit services remains in control of the services they use. Any accident, even involuntary, or happening without the knowledge of the credit user, negatively impacts the credit score. Rebuilding credit is a long and painful road. Our online reputation on social media uses a similar model and establishes how much people appreciate what we are doing. Online reputation rewards contradictory things: how socially acceptable and mainstream we behave, or how much outrageous and dramatic we are. Inside technology, online reputation is connected to our usefulness in helping others. Hiring factors the online reputation as an essential factor. Between two candidates with an equivalent skill level, the one who has the most visible online presence and highest reputation will be selected.

Sharing applications use reputation. Uber or Lyft drivers, AirBnB hosts are rated. They also rate their customers. When using these applications, any record is being shared. This has the positive effect of encourage friendly behavior when using these services, but also generate some average conformant, expected, average behavior. Consequently, it is strongly discouraged to use any of these services when in a bad mood, or stressed situation, because the consequences of any incident can be long-lasting. When cars will be able to detect, through sentiment analysis, that a person is too stressed out to be able to drive a car, the car won't start. As some one objected during a conference presenting this issue, what if a person tries to escape by car from a criminal having penetrated into their house? Or if you are trying to rush to a hospital with a family member having a heart attack or a stroke?

China is using credit scoring and pushing it to a whole new level. Social credit scores measure how much an individual supports the actions of their government. People who have poor credit scores, either because they are offline or they refuse to play the game, are cut off from several services, and are prevented from certain

essential services, like the ability to travel [^AW-China-Travel]. Friendliness, good reputation, conformance to social norms can gradually lead to obedience and servility in an authoritarian regime.

[^AW-China-Travel]:Alexis Wichowski, *The Information Trade — How Big Tech Conquers Countries, Challenges Our Rights, and Transforms Our World*, Harper Collins, 2020.

Loss of responsibility for content publishing makes rumors, fake news, political propaganda, technically indistinguishable from real, well-curated, respected source news, and it is practically impossible in practice believers that their sources are not to be trusted. The equality between truth and lies is poisoning the political atmosphere, creates extreme polarization and diminishes the credibility of any form of social contract, rights and duties, especially in democratic societies.

As everything we do is broadcast and reported, abuses happen also because of misinterpretation of what we do. A university professor who pronounces the N... word, even to analyze the way it has been historically used, is quoted having said it, and can be dismissed because of a clip taken out of context. The traceability results in people losing jobs.

Complete transparency, as provided by social media and made totally indestructible by Blockchain can be a good thing and at the same a bad thing. Restoring identity jeopardized by identity theft, escaping from bullying, getting on a no-fly list because of homonym situation, creates a feeling of despair.

Uselessness, indifference, passivity, despair, powerlessness, conformity, loss of common references, are among the side effects generated by technology. From there, it is tempting to consider that democracy is something from the past, and authoritarianism is the only viable solution for the future.

This resignation however negates two factors, that need to be analyzed, and could serve as a basis to fight against it.

1. Those who have vested interests in pursuing and aggravating the current trends know very well how to push it, they understand better what to do to extend their influence, and can do that without obstacles. It is important to

understand how they operate, and find what can be done to limit their ability to expand their grip. What have they understand than others have not? Is optimism equivalent to naïveté?

2. Technology is not a black box that can't be opened. It's a universe populated by lots of people, who play very different roles, it's made of lots of layers, it has become complex enough that it's almost impossible to grasp it into a single overall concept. When seen from inside, technology has many aspects. Many of those aspects are not hidden, they can be learned, studied, practiced. Many parts of technologies are open-sourced, and anybody who is curious can get acquainted with a piece of it.

This list contains 328 concepts, activities, domains, that are arbitrarily extracted from the domain of technology. Each of them is linked to a specific activity, with a specific set of approaches, tools, and perspectives that each support specific skill sets, and professional careers. This list is far from exhaustive, it has been created by looking mainly at the job offers in technology. Please note that this list doesn't contain any acronym nor any specific language or product, just concepts, methodologies, generic approaches and product types. It is only the tip of the iceberg of what technology is. However, many of the terms listed here have a direct impact on our daily lives, on the products we use. Here it is:

Access, Accuracy, Actionable insights, Agile, Algorithms, Application Programming Interface, Applications, Architecture, Artificial Intelligence, Assembly Language, Asynchronous, Attack, Audio, Auditing, Automation, Back End Development, Backend, Best practices, Big Data, BlockChain, Browser, Bugs, Business Intelligence, Business Value, Canvas, Cascading Style Sheets, Certification, Change Management, Channels, Chat, Client, Cloud, Cloud Development, Code Review, Coding, Collaboration, Commands, Commissioning, Components, Computer Science, Concept Maps, Confidentiality, Connected Applications, Connected Devices, Connections, Consulting, Containers, Continuous Deployment, Continuous Integration, Copyright, Credentials, Cross Site Forgery, Cross-functionality, Cryptocurrency, Cryptography, Customer Feedback, Customer Success, Cybercrimes, Cybersecurity, Dashboard, Data, Data Analysis, Data Bricks, Data Collection, Data Curation, Data Entry, Data Extraction, Data

Governance, Data Ingestion, Data Integration, Data Lake, Data Loading, Data Models, Data Operations Analysis, Data Processing, Data Science, Data Scientists, Data Structures, Data Transformation, Data Typing, Data Warehouse, Data-driven solutions, DataOps, Databases, Decommissioning, Defect Analysis, Deliverable, Delivery, Deployment, Desktop, DevOps, Diagnostic, Digital Applications, Digitization, Directories, Discussion Group, Distributed Denial-of-Service Attack, Documentation, E-Commerce Platforms, Extract-Transform-Load Pipelines, Elastic Search, Electronic Payment, Email, Encoding, End Users, Engineering, Enterprise Data Systems, Environment, Environment Variables, Evangelism, Face Recognition, Features, File, Files, Folders, Forensics, Format, Frame, Framework, Front End Development, Frontend, Full Stack, Full Stack Development, Functionalities, Garbage Collection, Geospatial Data, Global Network, Graph Database, Hacking, Handwriting Recognition, Hardware, Helpdesk, High Level Language, Hosting, Hypertext Markup Language, Information Technology, Information Technology Operations, Information Technology assets, Implementation, Industry Standard, Infinite Loops, Information architecture, Infrastructure, Instructions, Intellectual Property, Intelligent Analysis, Interfaces, Internet, Internet Service, Knowledge Graph, Knowledge Management, Knowledge Network, Languages, Laptop, Latency, Libraries, Linked Data, Local Storage, Logging, Login, Longitudinal dataset, Loops, Machine Learning, Maintenance, Maps, Mass mailing, Memory Leak, Messaging, Micro-services, Mind Maps, Mobile Apps, Mobile Apps, Mockup, Modeling, Modules, Natural Language Processing, Network, Network Engineering, Network Manager, NoCode, NoSQL, Object Database, Object Relation Mapper, Object-relational mapper, Office tools, Online ads, Open Source, Operating System, Operating System, Optimization, Orchestration, Organizers, Outages, Pair programming, Parallel processing, Penetration Testing, Performance, Phishing, Photo Editing, Pipeline, Platform as a Service, Platforms, Platforms, Port Scanning, Ports, Preventative Maintenance, Privacy, Problem Solving, Problem definition, Problem-solving, Procedures, Production, Production Systems, Programming, Programming languages, Project Management Software, Proprietary Software, Purple Teaming, Push / pull, Quality Assurance, Queries, Query languages, Rapid Development, Raw Data, Read-only access, Read-write access, Recommendation Engine, Relational Database, Representation of results, Resilient Technology, Reusability, Reverse Engineering, Robustness, Rollouts, Scaling, Schedulers,

Schemas, Scientific Apps, Scientific Data, Scrum, Search Engine Optimization, Search Engines, Secrets, Security, Security, Semantic Web, Server, Service Deployment, Site Reliability, Slides, Smartphone, Social Media Platforms, Software, Software Development Lifecycle, Software as a Service, Sound Card, Specifications, Speed, Spreadsheet, Spreadsheets, Stakeholder Assistance, Standards, Startup, Statistical Analysis, Statistical modeling, Storage, Store Procedures, Streaming, Structured Data, Subject Matter Experts, Subscriptions, Summary Reports, Support Libraries, Support Specialist, Surveillance, Synchronous, System Administration, System Architecture, System Functionality, System Security, System Design, Tables, Teamwork, Technical Issues, Technical Support, Test Automation, Test Cases, Test Reports, Test Scrips, Testing, Testing, Text Processing, Threats, Topic Maps, Touch Screens, Traceability Matrices, Tracking, Training, Troubleshooting, Trust, URL, Unstructured Data, Updates, Upgrades, Usage Guide, User Experience, User Guide, User Interaction, User Requirements, User-Friendliness, Validation, Variables, Video, Video call, Virtual Private Network, Virtual Reality, Viruses, Visualization, Voice Recognition, Voice recognition, Vulnerability Management, Web, Web Apps, Web Apps, Web Framework, Webinar, Workflow, Workstations

Time to dive in.

This book is not a general introduction to technology that aims at explaining all the aspects listed above. But it is an attempt to start breaking the glass ceiling by looking at technology in what it is essentially: a human activity, made by many people. Time has come to stop considering technology as a secret, magical, god-like, impenetrable black box against which no human action is possible. In other words, if technology is an ocean, it's time to learn to swim, to avoid getting drawn into it. Not just as "users", but as citizens who have rights, opinions, and want be to stay in control of their lives.

Some of the obstacles are easy to overcome. As animals, humans tend to limit their territories by putting some unique signs preventing outsiders to get in. The technological community, as many corporations do also, has been prolifically using

acronyms. Acronyms are the best way to exclude newcomers from getting in. For example, these are some randomly extracted from job ads in Big Data: AWS, SAS, JPM, MATLAB, SQL, R, CI/CD, GIT, JIRA, GPA, HDFS, ORC, T-SQL, MDX, DAX, NoSQL, MECE, ETL. Whatever each of them means, whether they are products, languages, processes or methodologies, they mean that if you don't know them, you don't belong to that community, you are not qualified in those technologies, and you don't have a chance to get a job before you learn at least some of them, not only what they mean, but how to operate them. But that's comparable to any corporate or institutional jargon that is acquired by necessity. For example, 1040, 1099, W-2 are not just random numbers or combinations of letters and numbers. Anybody who is working or has worked in the United States understands that those numbers mean that, in order to file a tax return, you need to determine whether you have worked as an employee or an independent contractor and fill the appropriate forms.

Getting to understand not only the acronyms, but the concepts used in technology, is as essential as knowing how to appropriately file a tax return. Tax is governed by law. As such, it is possible to understand the bells and whistles of any legal provision, and it opens the ability for discussions and gives the ability to resist a government decision, if considered unfair. As Alexis Wichowski explains in her book The Information Trade, the net states are imposing their own rules, and there is no easy way for "citizen-users" to claim their rights, as they are corporations used to play by their own rules. Worst, when the net states principles are being implemented by nation-states, the citizens themselves lose their ability to oppose any decision. Their only choice is to follow the rules, and this is how authoritarian regimes are using technology as a powerful support for establishing a more and more absolute power.

Good uses of technology

The "net states", i.e. the Big Tech corporations who have gained a level of power and influence at least equal to the "nation states", inherit from the hope that technology will build a "better world", where access to knowledge, providing information immediately at their fingertips, for a small or negligible price, make a positive difference. Being able to be connected constantly can improve and save

lives. Technology usage is so pervasive and the rewards are so immediate that no user is willing to go back to a place before technology exists, regardless of its dark side effects.

Nefarious usages of technology

On the contrary, there is a lot that doesn't conform to that model.

It is possible to escape scrutiny and to transfer critical data over national boundaries to protect them from national laws that are constraining. Digital shelters are used to evade taxes.

The masters of political propaganda have found out that exploiting the borderless connections provided by social media was a powerful way to propagate extravagant rumors, to the effect of weakening trust in democratic institutions including the electoral system in other countries.

The authoritarians or would-be authoritarians have found out that the fine level of control enabled by technologies can be used to serve their purpose of controlling the opinions, and actions of citizens. The smartest ones use the social controls to reward compliant citizens. Instead of punishing the bad citizens, they found it was as efficient to exclude them from social goods or services now reserved to good citizens.

A mixed bag.

We are being told that you can not have the good side without the bad side. If we don't pay to access technology, it's because we accept to be tracked, and our information is valuable and serves the need of an important segment of the economy, allowing companies to target their customers in ways that are unprecedented.

However, things are evolving. We have reached a point where nefarious usages have become serious enough that it carries the dangers of commercial war (between China and the US), civil war (within the US), threatening the survival of democratic societies, free press, free elections, and the rise of authoritarian

regimes, partially inheriting from sinister traditions (fascism) but also quite innovative in the ways they have tamed technology usages to serve their needs.

There is no equivalent, on the democratic side, on how technology can serve the needs of the government. Or rather, there are attempts, in various levels of the governments, for example in cities, to try to reverse the trend and use technology to provide better services to the people. But technology remains a black box, that can't be penetrated. For example, campaigns organized by activists or proponents of new groups actively involved in changing things for the better have no problems using Facebook, the same company accused of playing a role in the demise of democratic values.

In a way, that's normal, as technology is neutral. The fact that we use the letters of the alphabet to communicate ideas, whether considered good or bad, is not held against the fact that we are using words made of letters. The fact that we're using digital information to communicate doesn't mean, similarly, that digital information and its associated technologies is good or bad.

The problem is, therefore, not technology itself. It's the way it is being used, the mechanisms that are at work. As it's easier to analyze how technology is misused than the way it's used, it's interesting to start digging about how technology is used as a political weapon to reach certain goals: political propaganda, fake news, rumors, bullying, human trafficking, tax evasion, money laundering, demeaning democratic values, weakening of institutions, including government.

It is also interesting to analyze which negative feelings are being generated by certain usages of technologies and how these results are achieved. Powerlessness, loss of control, uncertainty, loss of trust, blurring the relation between true and false, impotence, the feeling of overwhelming, impossibility to get a comprehensive picture, inevitability, resignation, deterministic belief that it can't be otherwise.

My hypothesis is that these results, far from being inevitable, are being engineered by people who know what they are doing, and are benefitting from the blindness of the other side, where their opponents are not able to comprehend or devise any strategy to counter it. We will study some of these results and reverse-engineer the

way they are produced, in order to give the means to those who want to counter them to address them.

To illustrate those points, it will be useful to pick some examples and dive deeper into the technologies that are used. There is no substitute to not getting into the nitty-gritty parts of the technologies used. Otherwise, things will stay as they are. On the one hand, technologists will continue to work at their micro-level to provide yet more efficient, more scalable, more pervasive, and less accountable technologies, not looking at whether their effects are positive or negative. On the other hand, people who benefit and people who suffer from the outcomes will continue to witness the growing gap between them.

Less is more, or how big can data grow?

Tracking tech users in a surveillance capitalistic society (S. Zuboff) has assured an unprecedented level of wealth and power to the net states (A. Wichowski) and at the same time has created an unprecedented amount of data to be processed. Technology has followed the trend, and the domain of "big data" has flourished. The results have been quite convincing and machines are now able to provide valuable, and targeted information, to the corporations that need it in order to better target their own customers.

The underestimated side effect of this trend is that the technology used to address smaller amounts of data has lost its appeal, and is considered old-fashioned. But humans are still humans, and the human brain, who remains, if we listen to neurobiologists, far much more sophisticated than any algorithm, tends to be considered inferior to machines, as there is a threshold to the amount of data it can absorb. For example, searching for, let's day, the term "artificial intelligence" in Google, from New York City, on December 11, 2021 at 1:30pm yields "About 767,000,000 results (0.59 seconds)". It's a remarkable achievement that Google is able to produce, in a fraction of a second, three quarters of a billion of hits for that term. But why would I, and any other human being, care about that number? This sounds ridiculous, but has the effect of showing how short our lifetime is, how limited our brain power, and it leaves us with the impression that although we want

to learn something, we'll never be able to get to the bottom of it, nor even scratch the surface. But it shows us also how great Google is, how performant it is, how efficient it is when the goal is to aggregate as much data as possible. In other words, it is self marketing for Google at the detriments of us, poor and powerless human beings.

This rant may end up sounding childish if it was limited to a pure assertion of how great Google is. It does have real world consequences. It encourages some people - inside corporations or governments - to assert that machines are superior to humans, in their efficiency to retrieve and process data. Human expertise is much more expensive (experts went to a long process of education), is less reliable (experts disagree sometimes), subject to failures (it may be biased), is much slower (it takes more than a fraction of a second for a human to retrieve relevant data on any topic).

And this leads sometimes to the replacement of humans by machines, in the name of progress, like horses have been replaced by cars, or messengers have been replaced by telephones, etc. However, what is absent from the assessment are the same questions that are being raised when dealing with humans. Is it that cheap (search algorithms are not 100% automated, there is a lot of human labor involved in the background)? Is this that reliable? People who create information and know their content often find out that Google fails to report what's most important on their content, because their algorithms are not tailored enough to address their unique content. Then, those who are not the information originators have no way to know how much valuable data they miss. It doesn't matter how many millions or billions of data points there are if the crucial information a user needs is missed. Bias also plays a role, because the results are computed by algorithms that process information according to certain hypotheses. These hypotheses are crafted by human beings, as well. Biases are much aggravated in automated processes. If a wrong, or incomplete, hypothesis has been used, a multiplier effect on the number of data amplified dramatically the effects. Then, like in the human expert case, how do we know if something is right or wrong? It may depend simply on what concepts have been used to calculate it.

Google makes us believe that the rules it applies are transparent by publishing guidelines for search engine optimization. These rules are quite detailed, and a company that wants to comply with us finds itself with procedures that are as arcane and complex than filing financial reports, but depends on Google instead of regulatory procedures created by accountable government bodies. As the search algorithms are kept secret, there is no way to know what other rules Google applies in order to rank search results and have them appear the way they are. A company's success or failure may depend on how it is ranked on search algorithms, and how it compares to its competitors. But there is no way for a company, even less for an individual, to recourse if the outcome is considered unfair.

There are exceptional cases where Google itself modified its algorithms for a cause they judge morally valid. For example, as Alexis Wichowski points out, searching for ISIS recruitment was diverted to sites that were discouraging it, and leading users to sites providing psychological and emotional support, as Google considered that this could lead to decrease enrollment. And it worked. In this case, the purpose is honorable and can be applauded enthusiastically, but this also shows that Google has the ability to orient the search results to anything it finds valuable at the time. If the winds term, and if these powerful tools fall into the wrong hands, the unattended consequences of such powerful tools is unfathomable. If democratic nation states have no way to weigh in to protect the freedom of expression while prohibiting certain behavior under codified laws, the consequences can be pretty dramatic.

Search engine technologies can be analyzed. The simplest technique is based on "string recognition". Either a word or expression is present or not. Then there are variants of it. Case variations are the simplest. Then there are variations that are more subtle but still useful, such as ignoring accented characters in search. Then there are grammatical variants: plural, conjugated verbs. The technology of natural language processing has been created to handle these cases. Then there is a semantic layer: if a name has synonyms, using either one of the terms could lead to the other. Generalizing this, taxonomies and ontologies have been developed to show each term in relation to others.

If a term has homonyms, then disambiguation techniques can be used. Specific search technologies has been developed for given domains, building on domain-

specific vocabularies.

Another aspect of search depends on whether the source data is structured or unstructured. Text is typically unstructured. A database is typically structured. Hybrid, semi-structured data sources are also available. For example, tags marking up text can document the nature of the portion of text within those tags, and therefore end up being equivalent to a structured information source.

All these techniques are used and provide ways for users to define their own research patterns. They can be modified and fine-tuned according to the expected results. Some of them are built on top of proprietary software, others are open-source. Companies provide domain specific vocabularies or taxonomies that can be applied directly on new information sources.

DuckDuckGo doesn't track users, but is their search engine open source? Google Search tracks its users, and its search engine is proprietary.

Clueless solutions

Climate Change Data Has Scientists Scrambling for Solutions.
https://www.wsj.com/articles/climate-change-data-deluge-has-scientists-

Solution found: "To speed access to the data and lower the cost of computer equipment, NASA and NOAA are working with Amazon Web Services, Google Cloud and Microsoft Corp. to move their climate databases into the cloud."

Chapter 2 Outperformed by Machines?

Artificial Intelligence / Machine Learning / Internet Of Things

Artificial intelligence is not new. The term was coined in 1956 and has been through several phases of development, with variable success. It has become a buzz word in the recent years and has attracted a lot of traction.

Why is that? The need to analyze enormous amounts of data available has created an incentive for improved automatic processes in order to produce aggregated results, that have a high value for marketing and sales, advertising, political campaigns and all forms of mass media. New techniques have been developed to increase speed and performance: microprocessors continue to be more powerful over time, and on the software side, new techniques are developed to improve performance. Parallel processing is the ability to use multiple computers in order to divide a task into smaller subtasks. Asynchronous programming is the ability to start processing data before all of it is loaded. Various techniques for optimizing data processing continue to be created and implemented. Caching enables to store and reuse data and only update the part of it that has changed.

Data is being analyzed using a variety of techniques, relying on algorithms for optimizing processing speed. An existing term was reused to describe the set of techniques and processes used to that effect: artificial intelligence. The term is somewhat a misnomer, as it is a set of human-created algorithms but the fact that they automate this workflow justifies the term of artificial intelligence. Intelligence in this context is to be understood in the sense of data collection and analysis (as in the Central Intelligence Agency), not directly related to the human feature of intelligence as opposed to stupidity. However, the term is being used as an argument that aims at proving that machines are better than humans as in "artificial intelligence is more powerful than human intelligence", as it handles more data.

Neural networks are built on the metaphor of the computer considered as a human brain. Neural networks are useful, but they pale in comparison with the capabilities of the human brain, especially when applied to the resolution of complex problems.

The research activities in neurobiology using computer models of the human brain still remain at a very early stage and the capacity of the human brain to deal with complex situations is unmatched, and this probably be stay like this for many years to come, despite the spectacular demonstrations of the machine capabilities.

In the general public, unaware of these nuances, the idea that computers are better than humans is making its way. It is clear that automation of repetitive tasks is clearly winning. This affects many more professional sectors than previously thought, including so-called intellectual professions: medical professionals, accountants, lawyers, experts from different fields benefit from help they received from computer-driven systems to collect, sort, organize and present data in ways that were time-consuming and tedious. Low-level tasks, such as boilerplate-based standard legal contracts, diagnosis based on comparing patient data with statistical data, are examples of tasks where automation has clearly benefitted its users, and may have resulted in removing low-level jobs. But the public perception goes broader than that and there is a feeling that many professions are impact, professional careers are threatened, and that the "machines are winning". This results in a loss of purpose, a sentiment that things are now switching to technology-only processes, that without technology nothing is possible. What is being blurred is an understanding of the processes that the machines are better at, some of them having never been done by humans. The notion that machine processing can be seen as a complement to human activity instead of a replacement is very marginal.

Machine Learning and loss of human control

Algorithms which are based on switching on and off zeros and ones at a very high speed, are repetitive. Given the same data set, an algorithm is doing the same thing over and over. There is no surprise to be expected. This is both a good thing and a bad thing. It's a good thing, because algorithms can be used many times, and, when they produce something which is different from what's expected, this often is interpreted as the sign that something is wrong in the algorithm, that there is some bug that needs to be fixed. It's a bad thing, because as data evolves, the algorithms keep repeating over and over the same things. If there is a need to fine tune

manually the results, this task needs to be repeated over and over each time the algorithm is run.

To improve the situation, the idea is to provide a way for the algorithm to analyze the data before it's done, and based on what it has found, modify the end results accordingly. For example, in a search algorithm where many people regularly mistype a word, the spelling error can be recorded and assigned to the proper word, so that the subsequent requests with the ill spelled words would yield the same results as if it would have been typed without the spelling error. This process of enriching the algorithms with more data that improve their efficiency at future runs is called machine learning. Machine learning algorithms are based on mathematical and statistical techniques, which are based on complex processes, and use various models, such as supervised learning and unsupervised learning.

The tools that implement machine learning operate as "black boxes" that take some data as input and produce other data as output. It is not always necessary when implementing a machine-learning engine to master all intermediary steps requiring a deep mathematical and statistical background. For the vast majority of the end users of such products, all these layers are not interesting, or too complex. People tend to rely on the results where they can see visible improvements without having to dig into too many details.

As algorithms can be programmed to modify themselves using these techniques, it comes a point where no human being, including the programmers who started the artificial intelligence project, have the ability to understand what is happening. They are able to tweak slightly the algorithms in the hope to get better results in the cases where it's needed, but it tends to become a guess game, with a whacamole effect. Tweaking algorithms is not easy, as they may produce side effects. The more algorithms are self-correcting, the more difficult it becomes to modify them.

Machine-learning based systems can be a real asset to help digging into huge sets of data, but if they are given the power to take decisions without the ability to check what is happening inside in detail, it can also lead to disastrous consequences in some cases.

The public perception of machine-learning technologies is that humans are becoming more and more irrelevant. Not only machines do -- or seem to do -- a more efficient jobs than humans, but there is a point beyond which they can't be tamed any more. They have their own logic, and all bets are off.

For those who still consider they are in charge, and are liable for any results that don't fit the expectations, the consequences can be devastating. Before that happens, they understand that they don't have the ability any more to interfere with the technology used. This situation can generate quite an amount of stress, and the sentiment that they lost control.

Losing privacy

Android phones report their location even when they're off (to be checked). For people who are trying to save their freedom to go where they please, and not having to report to any one, this can be a difficult thing to accept. There has been a slow evolution in the renouncement to privacy. Credit cards statements record any purchase we are making, web browsers and cookies leave a trace of our Internet activities, phones track every location we are going. At the same time, the convenience they offer makes it quasi impossible to live without. Not wanting to leave traces is suspicious. For example, trying to buy a plane ticket with cash raises alert levels. Never going on the Internet is in today's world unthinkable in practical terms. There are many services that are not even offered off-line, and it doesn't look like this is going backward.

The most frequent attitude about it is simply to ignore it and not think about it. This is fine as long as we don't do anything that can get us in trouble with the justice system. But in countries where you can be in trouble because of factors that don't depend of what you do, but simply who are: your religion, skin color, income level, social credit score (in China), political opinions (China, Russia, Turkey, etc.), then you can get in real trouble quite soon, with no real way to counter it. It is what is it, and you are left on your own, as in previous 20th century dictatorships, you were constantly followed and all your activities were reported. Your email content is open,

the materials you read online, etc., nothing is left in the dark. Freedom and privacy are much more connected than most people think, especially in America where so far the prosecution of people based on their political opinions has been limited, although there have been some dark periods where personal freedom was at stake.

For those of us who were, directly or indirectly, confronted to political or racial persecution, these restrictions to privacy rights are disheartening. Again, there is a feeling that this goes beyond our control. Regardless how much we pay attention to preserve our individual rights to do whatever we want, we know that we have lost control over it in a significant way.

Technology companies claim that an individual has control about their privacy settings. But the reality is murkier. There are plenty of situations where these settings are ineffective and don't provide real protections. The fine print is where this is hidden, and practically nobody ever reads the terms of service on the Internet. If we would start to read them, it's hard to agree with the majority of them and we wouldn't accept most of them. Not only we lose a big chunk of our individual freedoms, but we feel powerless because there is practically nothing we can do about it. This feeling erodes the trust into a democratic society, which is in principle governed by the will of the people. Where is democracy in an environment where people have lost their ability to express their wills?

So far, hospital patients are being monitored and any indication of a sign going in the wrong direction triggers intervention from a nurse or a doctor. Wearable medical devices, health apps, or even "smart beds" are able to constantly monitor our health condition, and have the potential to save lives by signaling early signs of some vital disfunction. These signals can be sent wirelessly and trigger a quick immediate response. This looks like a good thing to have. But the price to pay is to accept that constant monitoring and therefore lose any sense of privacy, as even in our most intimate places - the bedroom, potentially also the bathroom - data is collected about our gestures, heart beats, and other vital signs.

This is the individual equivalent to the societal change in which societies went through by reinforcing surveillance, therefore limiting freedoms, in the wake of 9/11 terrorist acts. The vaccine mandates are similarly decisions taken by the

government to protect the population by imposing constraints on the individual right to choose whether to be vaccinated or not. In order to protect the rights of the unborn, some states decide to prevent women from terminating pregnancies. In order to protect the life of their citizens, some states have strong restrictions on gun ownership while others have not. Each of these decisions is controversial, and disagreements are based on various perspectives on moral and religious values, the role of government, the balance between individual freedom and the feeling of belonging in a society as a whole. There is no clear-cut answer to any of these questions. A democratic society is characterized by an ongoing public discussion on these issues, and the ability to express opinions, and to regroup with political allies to change things through public debate and voting. A true democratic society is when people have a voice, and the majority of the people can decide. It remains to be seen if the United States, in that respect, still can be counted as a democracy.

Anything made possible by technology is not automatically acceptable without discussion, just because it's possible. When it creates a potential for controversy, it should be submitted to public debates. For example, the ways by which connected products (the Internet of Things) report data should be debatable, and consumers should be fully informed about what a product is capable of reporting before deciding on a purchase. Abuses can result from providing incomplete information. Consumer protection is more and more connected to privacy abuses. It is clear that the price to pay for a democracy to survive is to reign in the surveillance that needs to be accepted through a clear agreement. Having to go through the TSA checks before boarding a plane is clearly a collective decision that depends on the political power and the government. Having one's heartbeats monitored while sleeping should be an individual decision. As the ability to terminate a pregnancy. Vaccine mandates and the right to carry guns is of a different nature, as it endangers not only targeted victims, but also neighbors and bystanders. The parallel with driving is that the mandate to respect traffic lights is of a collective nature, whereas the mandate to fasten one's seatbelt is of an individual nature. However car manufacturers have the obligation to provide cars equipped with seatbelts.

A new right to share should be part of an updated bill of rights. Companies could compete on what they share. If a service is more expensive because it doesn't

share, let's be it. Many business offer paid subscriptions with no ads. As data sharing's principal purpose is to serve the advertisement industry, bailing out should be an option, and many people would gladly pay for it, if offered the option. This extra charge could be called a stamp. The Post office carries letters by requested a fee paid through a stamp. In exchange, the post office guarantees that the mail doesn't get opened. There is no reason why email providers could not guarantee that they are carrying the messages without opening it, even if it goes through a subscription service. Google offers professional email service for a fee, but still reads the content and resells it to its clients.

Chapter 3 Inadequate Legal System

The Internet is an open infrastructure, by design. Therefore, restricting its access is very hard, and the basic restrictions are often mere band-aids.

Going on the Internet is like walking on a street. It's very hard to remain anonymous, and not being recognized. We would need to get to ridiculous ways of hiding our face and appearance, to avoid being recognized. Perhaps the combination of a burka, a face mask and sunglasses may prevent others to identify us.

Providing login credentials to enter a private web site is similar to showing an ID when entering a building. It is not 100% secure. It is likely that showing an ID of a person whose photo looks similar to you would get probably get you in most times. Computers are more strict, but bots are able to steal login credentials or in some cases try a multitude of combinations until one works. New protections are working better, such as 2-step identification, including one that sends text messages to your phone. If you do not have your own phone near you, then you can't go further. Of course, this method would not protect you if you lose your phone or it was stolen from you.

Protections are like dams. The DDOS attacks are the equivalent of flooding. The number of requests on a server get so big that the server is not able to offer resistance against the number of simultaneous attacks.

Software has vulnerabilities. Even proprietary software can be hacked, and hackers have developed ways to find those vulnerabilities and to penetrate systems. This is like finding loopholes and using trojan horse tactics. Once you're in, you can go anywhere.

Most users feel they are guilty to not be obedient enough to the most recent rules set to protect them. For example, a 2-step identification is more secure than 1-step. Using a smartphone as a physical device that can only be in the possession of one person at a given time is considered to be a good security measure. If you can prove that you're close to your phone, and you can click "yes" on it, it proves it's you.

Overall, cybersecurity remains an oxymoron. Again, being secure on the Internet is as likely as walking in a street and not being recognized by anyone, including your next-door neighbor. If you don't want to be seen, don't go out, that will fix it. Moreover, make sure that no recording device is present that could report where you are at a given time. Eventually, this amounts to: if you don't want to be seen on the Internet, don't use any device that is connected to the Internet. That goal, in this day and age, is very hard to achieve.

Private networks, such as Intranet, which are reserved to exclusive use of an organization, use the same techniques than the Internet, but strictly limit access to authorized users. However, this hybrid approach is not entirely secure as it still uses the same technologies and software that the ones that are massively available on the open Internet.

Governments, and organizations which need to ensure full security in their communications, use alternative networks. They are quite exclusive, as they are very expensive to build, operate and maintain. The security challenges they face have more to do with the trust of the people involved. The security protocols under which they operate have to be extremely strict, much stricter than the ones used on the open Internet.

However, on the Internet at least, hackers are getting more and more sophisticated and find creative ways to penetrate servers that are supposedly well protected. In

other words, absolute protection is very hard, and the most secure environments are those which are not connected through public networks such as the Internet.

As critical infrastructure depends on Internet-based technologies, vulnerabilities grow and the consequences of the threats can be deadly. Cyberwar is not any more in the domain of science fiction. It is as concerning today, than the danger of nuclear was in the late 1950s.

Everybody can publish

Publishing on the Internet is very easy thanks to social media platforms. The rules for publishing are much looser than the ones that are needed to publish print materials. With print, one needs to create, print, and distribute using heavy technologies such as printing presses and efficient distribution channels. With an account on a social media platform, within a few minutes it is possible to become a publisher.

Freedom of speech is protected by the first amendment right, and equivalent rules and statutes in democratic societies. In not-so-democratic societies, rules are more stringent. Publishers have some responsibilities and duties and are liable for their content. Some legal provisions enable those who feel their integrity or reputation have been attacked to sue publishers. The rules against abuses committed by individuals publishing online on privately owned platforms are murkier. Bullying, propagation of false rumors and fake news, is therefore much easier. As publishing platforms are looking for content that keep their audience riveted, they are less likely to stop outrageous content. In the printed press world, we know what to expect when reading a tabloid. If we are not interested by this type of content, we can choose to read a more serious newspaper. Reputations are at the publication level, because for every publication, the publisher takes responsibility over its content. On social media platforms, all content is published in a uniform form, there is no decentralized distribution of editorial responsibility, and anything goes, without the ability to distinguish clearly what is based on proven facts from what is based on political propaganda or false rumors.

Once any data is published, it is very hard to "unpublish", i.e. to remove it from public view. It's as complex as trying to get the toothpaste back in the tube. This can be very distressing if an unauthorized picture or video is published accidentally or by another person in an act of malfeasance. There are many stories of young people being bullied online, with tragic consequences.

Cryptocurrencies

Cryptocurrencies have been designed to circumvent the traditional financial system, and can be used both to facilitate money exchanges worldwide, for good or bad purposes.

Cybercrimes are on the rise, and are using several loopholes made available by the creation of cryptocurrencies.

Losing control over legislation and distribution of information

The provisions of the existing legal systems are not adapted to the online world:

- national laws can be circumvented by the borderless, world-wide circulation of information over the Internet.
- in federated countries, such as the USA, the state-based commerce laws are not adapted to the interstate provision of services enabled by the Internet. Any product that is available online is subject to interstate commerce, and this makes the handling of commercial laws very hard to abide to.
- social media platforms publish content without bearing direct responsibility for the quality of the content. Therefore they are less accountable than traditional publishers, and their liabilities are dissolved. The broadband companies, on the other hand, influence content as they attempt to remove the net neutrality features. It's similar as if a train or truck company would decide to privilege who gets to be privileged in receiving printed materials depending on how influent the receivers are. This lives us in a pickle, as a society.

Losing information.

Online information is not easy to preserve. Some of it has only a transient existence. There is no archiving available equivalent to libraries preserving books. There is no central repository in which every online material is listed, as is the case for the Copyright/ISBN office assigning a unique number to every published material, so that it can be retrieved.

Chapter 4. Losing Control and Initiative at Work

Computers are everywhere. For many people working means operating software on a computer, collecting and analyzing data, producing new data, communicating information, etc. That goes beyond traditional white collar occupations, as manual workers also use tablets or digital devices on their workplace.

Overall, digitization means that work is easier. Repetitive tasks are less frequent, overall organization is more rationalized, and working conditions have improved.

However, some drawbacks have also started to emerge. Because workflows are digitized, they therefore can't be easily modified without interfering with the software. This can lead to more rigidity, or sometimes the mere impossibility to resolve a problem because "the computer system won't let us do it."

Many issues have to do with customer support. First, the voice mail systems are often irritating, because they are designed to give the illusion that we are talking to a person, whereas we are talking to a computer. The number of options is limited, it often doesn't include the exact question we need an answer to. And if the computer doesn't understand the question, it keeps repeating over and over the same thing. As the computer speaks with a human voice, it's like talking to a person who is really really stupid. Not only customers don't get what they are looking for, but it

gives a bad reputation to the company that seems not to care about their satisfaction.

Once we get over this first obstacle, and we are able to speak with a human being, we may end up explaining the problem, and getting the answer: I understand your concern, but our computer system doesn't allow us to do what you want. If it were me, I would have done it, but I can't.

There we have an admission, that working people have lost control over the ability to take decisions because they are subject to rules that depend on the technology and on which they have no weight. They have lost control over what they can do, and they accept it as something inevitable. They expect the caller to understand it -- implicitly recognizing that we are all in the same boat, subject to the powers of an unreachable, higher level, technology under.

They could have given a different answer: we understand that currently our computer system doesn't allow us to do what you need. We will work on it to ensure that in the future your expectations will be met. In doing so, they would recognize that technology is evolving, that users have their word in how it evolves.

But the fact that it's not happening is because end users (here the company providing services to their customers) have lost hope that they can do something. There is an invisible glass ceiling which is considered unbreakable, and that amounts to consider that anything technical is out of reach.

Technology is complex. Naïve and uneducated users think they can do what they consider simple things with computers, only to discover soon enough that they don't get what they are asking for. There are many layers involved, they keep evolving, and just making something work at all is extremely challenging. This explains why the software industry is getting more and more concentrated into the hands of a smaller number of companies.

More experienced users are painfully aware that they won't get everything they want, and whether they have unsuccessfully tried or not, they realized that they are better off adjusting to what they get rather than asking for something that doesn't exist, even if this is what they would want. Ironically, "what you see is what you get"

(WYSIWYG) was the way to sell progress in the 1980s showing that a page could be tailored at will before getting printed. Now the same expression can still be used and understood as: "if you don't see it, you don't get it"!

More than one paradigm

Big data rules the world. That proposition holds for advertising. In order to find where your customers are, it's convenient to analyze as many data points as possible, run all possible filters to target a precise segment of the market, so that you can get the result you need.

Big data needs have generated big data technologies. Algorithms have been adjusted to scale up, optimization techniques are applied to accelerate the production of results. The mere size of the data available has generated a set of new tools, techniques, and technologies, that are able to produce impressive results handling an impressive amount of data.

Because of the importance of that market sector, a lot of investment has been made in this domain, and the Big Tech companies have created unique technologies that are used to this goal. And because of all this accumulated trove, they have been eager to reuse the same tools and techniques all over, and have propagated the idea that using technology means using big data technology. For example, the job interviews for tech jobs are based on "data structures" and "algorithms" created for this big data ecosystem.

However, this obfuscates the fact that many technology users are dealing with data sets that are much smaller (even if at their scale it seems big). For example, a government agency produces data, that is reasonably large, but would consider to be ridiculously small compared to the scale of big data. Instead of tera- or peta-bytes of data (big) we are are talking here of a few hundred thousand of data points. Although this seems overwhelmingly large, it is small compared to the scale of big data.

Any technology is not always the best choice at any scale. The "small data" world has several distinctive characteristics that differ from the "big data" world. The

biggest difference is that in big data, nobody is interested to look at individual data points, other than at an aggregation of data points considered at a statistical level. In small data, for example in a government agency, it is possible to find who has created a subset of this data, and ask them why these data points have been designed the way they were. If there are adjustments to be made, it is possible to discuss them. Also, there is a requirement for an agency to disseminate information which is entirely reliable, meaning not just 95% correct, but should be as close as possible to 100% correct. The reason for that is that the people who created, and/or published that information are accountable, they bear the responsibility to get information that can be considered as an authoritative reference in the matter, and therefore have no choice but have to ensure that their information is entirely dependable.

These are two completely different use cases, and it should come to no surprise that the kinds of technologies to produce data in either case should be different. The reality is that they are not so much different. The weight of the big tech companies involved in big data software and services is so disproportionately high that they make everybody believe that their solutions can be applied even when they obviously would not, if we think just a little bit about it. But thinking about it is not what people are used to, unfortunately.

Losing Control

Not only people lose control about what they can do to better serve their own needs, but they also lose control about what tools they use in order to satisfy their requirements. Often, they don't even know that alternatives are envisionable, and they are resigned to the idea to what they are looking for is unrealistic. This creates a feeling of passivity and uselessness. If they know that other solutions exist but can't get them, it created a feeling of frustration and powerlessness.

Either way, the outcome is not great!

Chapter 5

Can't compete against Big Tech monopolies

5- Losing Control inside Tech: Amazon vs. Elastic Search. Technology evolves rapidly. Graphic Design vs. User Experience vs. Javascript. Engineering, Open Source, Hacking, Code of Conduct. Inadequate Job Hiring profiles. Role of algorithms and data structures. No No No: NoCode, NoSQL. The Language War/ Operating System War. Is it still relevant? - Tableau, Miro, Airtable, Notion, Zoom, etc. The new business models. Where have IBM and Microsoft gone? Platforms. Frontend, Databases, Frameworks. The role of Facebook, Google, Twitter, Uber

Brief history

Today's technology can be divided into three main subdomains: devices, networks and platforms. When personal computers were first released, they were productivity tools aiming at providing better ways to perform office work.

In order to better understand technology, lets divide it into several domains.

- Devices: Computers, tablets, phones, IOTs, watches, glasses, digital assistants. They provide connection to the Internet.
- Networks: Broadband, Servers, Internet, World-Wide Web, Intranet, Cloud, Private Virtual Network. Orchestration, Automation, Servers.
- Platforms, Software, Operating Systems, Languages and Ecosystems: Facebook (including Instagram and WhatsApp), Google (including Android and

YouTube), Amazon, Microsoft (including Windows, Skype, LinkedIn), Twitter, Zoom, Uber, AirBnB, Apple (IOS, MacOS), Open Source, Oracle

1980s: Computers become personal

- Main players: IBM and Microsoft
- Important outsider: Apple

1990s: WWW

- Silicon Valley is the center of the world.
- New disruptive technology about the change the world.
- Online Media and E-Commerce

2000s: Collecting data for the advertisement industry

- Google does garbage collection and finds treasures, that they can sell.
- Facebook baits people into connecting with each other and sells their data
- Big data starts.

2010s: Smartphones + Cloud + Bad actors enter into play

- The whole world gets digitized through smartphones.
- Industrial Concentration in the US, unregulated.
- Russians see tech as an easy way to extend their influence into propaganda
- China expands its role in producing tech, and transforms its society into tech-driven dictatorship

Technological evolutions

- Predominance of big data technologies.
- How big tech open-sources part of their technologies to create satellites. Google-> Angular + Material, Facebook -> React, Twitter -> Bootstrap, Uber -> Vis, etc.
- Government and universities get on dead ends: The graveyard of good will thinking: Upper-level ontologies, taxonomies, Semantic Web, XML, etc.
- The complexification and multiple divisions of technological layers
- Big eats Small: Amazon vs. ElasticSearch, etc.
- The big tech propaganda main lines.
- Evolution of software trends:
    - Databases. SQL / NoSQL with query languages.
- Resilience / Resistance?
    - The Hacker Community, Open-Source, Public-Interest Tech, City-supported tech (NYC), etc.
    - Independent frameworks: E.g. Vue
    - New Programming Languages and Paradigms.

Control Lost:

- Impossible to function without a smartphone. Too many services have become indispensable.
- Impossible to develop software without the cloud: operating-systems based systems are decaying. A computer becomes (again) a terminal. Impossible to disconnect from the web.

Silicon Valley has stopped to be an innovation engine

Back in the 1990s, the Silicon Valley was in ebullition. The advent of the World-Wide Web was going to change the world for the better, disrupt most of human activities. Since then, disruption has happened, things have been disrupted. Lots of good things have happened (as forecast) but also lots of bad things have happened (first dismissed as marginal, then increasingly visible).

Big Tech has concentrated the intellectual and innovative workforce. Innovations are now limited to what serves their business model. Other forms of innovative thinking, especially if they result in making their business model in danger, are ignored, or killed. Universities and public research on technology is not on par with big tech research departments. Big Tech uses academic institutions to recruit for themselves, and have a paternalistic, "charity-like" condescendent behavior. They know they can't be beaten at the technical levels, because they hold the keys to the complex layers of technology involved. The cost of entry into tech-based research has dramatically increased.

Chapter 6

Towards an NIT

An unfulfilled need.

Social media platforms have become immensely influential; computers process unprecedented amounts of data; and the increasing complexity of Information Technology is getting more and more complex, and follows its own rules and structures that are impenetrable to most everyone.

Big Tech companies and potential bad actors in government have little account-ability when deploying such systems—in light of secret, proprietary algorithms. Similarly, heavy-handed (let alone authoritarian) governments can exploit these

technologies to blur the differences between truths and falsehoods. Russia and China are examples—even to the extent of interfering in democratic elections overseas.

A democratic society such as the US needs to have a clearer vision of how it can tame technological evolution to simultaneously foster innovation and preserve the rights of its citizens against authoritarianism, abusive attacks against privacies, monopolies, etc.

A precedent: NIH

- The National Institute of Health provides leading-edge research and is considered the best medical research center world-wide. Not only it develops its own research and product activities, but is the source of funding for the domain of biological and medical research for academic and private-based research activities. It works in close contact with the agencies such as CDC, FDA that regulate the market, authorize new products.

The tension between net states and the US nation state

- Non-ingerence deals (see Zuboff on Google with CIA after 9/11 and Google and the Obama Compaign)

The political need for digital literacy

- The congressional hearings with Zuckerberg were not only ridiculous, they are outrageous.
- Breaking monopolies is not the appropriate answer to nation-state vs. net state
- Democracy needs to be preserved. The nation-state can't dictate everything.
- Big Tech corporations should be regulated.
- New laws should be created to preserve individual freedoms.

The need to design nation-wide tech projects

- The nation-state can become a competitor against big tech, for example by providing a mail service that respects privacies. A national email service could be created based on the post office mail system, i.e. that respects privacy: as letters are not opened by the Post office, a national email service could guarantee that the contents of the email can be confidential. Only judicially-permitted investigations would justify to open the mail by law. In other words, any law-abiding citizen would be able to exchange email privately.
- A public search engine could provide access to knowledge without being tracked. Public libraries traditionally preserve the right to their patrons to read any books without recording what they read. The current universal tracking system doesn't protect the right to access information needed. For example, it is currently risky for anybody to find information about terrorism, even if the only goal is to understand how it works, where it comes from, in order to be better informed. Searching for information about terrorism carries the risk to be flagged as a would-be terrorist.
- A public social media could be created, that would have the ability to connect people together and would do that explicitly without tracking. This would foster communication between people without them being inundated with ads. This public social media network should be well protected to preserve it to being surveilled either for marketing or political purposes. Abuses should be monitored, and new laws should be created to prevent it from happening, and give a recourse to citizens who claim that their privacy rights have been infringed.
- Voting machines should have their software being disclosed as open source, allowing tech savvy citizens to analyze them, find potential bugs and loopholes, and there should be an open discussion about their maintenance. Only this publicly available software should be authorized in any machine that is being used for voting, at any governmental level: federal, state, municipal. This would go a long way to restore public confidence in the electoral system.

- New legal provisions should be studied and experimented to provide a publishing system adequate to the digital age. Publishers should be able to publish freely, in accordance to the First Amendment, with some possible limitations regarding hate crime, support for violent actions. Publishers, including individuals publishing on social media, should comply with those provisions and should be held liable in case of violations. This model should be created by looking at the best practices from the news, media, and book publishing worlds in a democratic society. Publication of information with detrimental effects, for example harrassment and bullying, should be compared against existing libel laws. New specific laws should be enacted, and find the balance between freedom of speech and respect of individual rights and reputation. In any case, the level of responsibility should be possible to establish. Broadband service providers, social media platforms, and creators of the information should have their responsabilities defined with the existing, or new, legal provisions.

- Secure cloud services and storage could be provided that would prevent any abuse, such as the Amazon vs. ElasticSearch story, in which Amazon, who is the cloud providers, has "stolen" the information stored by ElasticSearch and has created a competing technology that has jeopardized the commercial success of its client. Again, government-supported cloud service should guarantee that the information remains well protected. Obviously, the terms of service should prevent all kinds of nefarious data to be stored on those services.
- Research on accountability of information systems. This not only affects the financial sector, with projects involving cryptocurrencies and BlockChain, but also research on designing ways to dig into artificial intelligence and machine learning algorithms and reverse engineer some of the processes in order to demystify the magic that these systems have, with possibly negative consequences. More specifically, the ability to understand and investigate how these systems work could be of great benefit for many.
- A secundary benefit is that once the government is able to create its own, technologically advanced products, it will make it easier to trick down this information to Congress and all elected officials and get them a much more

precise knowledge of what technology is, where it is heading, and what they can do about it.

- It happens that benefits of technological advances are for some time disconnected from short term profits. The investor-based model driving the funding of the start-up ecosystem is looking for immediate returns of investiment, therefore does not support any project that is incapable of showing them. The intense research activities inside big tech corporations is aimed at reinforcing their business models, and therefore neglects other projects that don't fall into that framework. Therefore, there are a lot of projects that don't get funded, that could lead to breakthroughs in the future. Only government-funded research could fulfill that need. It is essential for US competitiveness in the world that it stays on top of the technology advances, because the current trend is that the authoritarian regimes are getting a solid grip on technological advances while democratic societies are letting the private sector lead and are taken aback consequently, as they lose control over the directions in which technology is heading.
- Some public-interest projects could reverberate world-wide and restore American influence. For example, if some technological breakthroughs can be used for better monitoring climate change, reduce inequities, and increase the general public literacy towards technology, there will be incentives to export this knowledge to other countries and foster cooperation, education, etc.

References

Texas University Press

Web site: https://www.ttupress.org/books/ Nothing on technology

Cambridge University Press

Web page: https://www.bookdepository.com/category/1708/Technology-Engineering

Book to read: Runaway Technology: Can Law Keep Up?, Joshua A. T. Fairfield

In an era of corporate surveillance, artificial intelligence, deep fakes, genetic modification, automation, and more, law often seems to take a back seat to rampant technological change. To listen to Silicon Valley barons, there's nothing any of us can do about it. In this riveting work, Joshua A. T. Fairfield calls their bluff. He provides a fresh look at law, at what it actually is, how it works, and how we can create the kind of laws that help humans thrive in the face of technological change. He shows that law can keep up with technology because law is a kind of technology - a social technology built by humans out of cooperative fictions like firms, nations, and money. However, to secure the benefits of changing technology for all of us, we need a new kind of law, one that reflects our evolving understanding of how humans use language to cooperate.

References

Social Media Alternative

Look at rova.app (Daniel)

Positive, Curated Content

Charles Thorburn: https://www.javelin-beta.one

Trust

- https://www.edelman.com/trust/2022-trust-barometer
- https://www.nytimes.com/2021/12/10/opinion/scams-trust-institutions.html

Rogan

- https://cdn.substack.com/image/fetch/f_auto,q_auto:good,fl_progressive:steep/https%3A%2F%2Fbucketeer-e05bbc84-baa3-437e-9518-adb32be77984.s3.amazonaws.com%2Fpublic%2Fimages%2F0e198b30-53a3-4316-96d3-209446172475_930x1102.png

How I escaped the Alt-Right

- https://www.youtube.com/watch?v=94_5mXsQTpA

Institutions

- Public Knowledge: https://publicknowledge.org/. Contact: Chris Lewis
- Creative Commons: https://creativecommons.org/
- Shorenstein Center (Harvard): https://shorensteincenter.org/. Contact: April Glaser
- Tech Lab, Harvard: https://techlab.org
- Yale Information Project: https://law.yale.edu/isp/
- Joint Center for Political and Economical Studies: https://jointcenter.org/
- Center for Humane Technology: https://www.humanetech.com/

Bibliography

Technology Science, Harvard: https://techscience.org

Shorenstein Center

Upside Down

Democratech

Why is there no government-funded research on technology in the United States? Why is technology treated differently from life sciences, where government-funded research through the National Institute of Health provides a clear competitive advantage worldwide?

Why are authoritarian regimes able to reign in technology companies and exert power over what they can and cannot do and while at the same in the US there is a sense of inevitability that technology leads us in a direction that is not controllable? Is control over technology synonym to dictatorships? Are European democracies limiting freedom of speech with the privacy requirements (GDPR and the like)?

Can individuals and citizens raise their voice about what is being done with their own, individual, private data? Is privacy a feature of the Luddites?

Why is the Post Office carrying mail by preserving privacy of individual correspondences while email carriers have the ability to read what people write to each other? (Same for phone, text, and chat conversations).

Why are book publishers liable for the content they choose to publish while social media platforms decline to take responsibility for the content they propagate?

Why are artificial intelligence providers

The Big Unknown

Two scenarios:

1. Democracy dies

- overfishing model.
- economy for the rich
- elections don't count
- people in power stay in power
- technology is the hands of the private corporations, and they can do as they please, provided they give to the government the surveillance tools it needs to suppress opposition.
    - no national cloud infrastructure. govt depends on Amazon, Microsoft for cloud infrastructure.
    - no national mail service
    - no national search engine
    - no national social media platform.
- accountability, liability, the rules of law become irrelevant.
- civil war done by cyber

Congress only scratches the surface

Congress is attacking Big Tech by claiming companies are acting as monopolies. For example, they don't provide equal access to their competitors in their app store[^politico-2022-01-04]. Big Tech companies are mobilizing their users to support them, by claiming that they may have to rescind popular services: Amazon said the proposed legislation could force the shutdown of third-party marketplace. Google claims that the bill will break popular consumer services while making less private, less safe and less secure, and they could be forced to remove businesses information from its maps and search results. [^politico-2022-01-04]: Emily Birnbaum, "Amazon and Google deploy their armies to twart antitrust bills", *Politico*,

January 4, 2022. https://www.politico.com/news/2022/01/04/amazon-google-thwart-antitrust-bills-526460

What this story shows is that Big Tech companies are not ready for any change to their business models, even as tiny as the one proposed by Congress. They have successfully convinced a significant portion of their users that their success depend on them staying in power. This argument resembles to blackmail, -- if we fail you fail. Worst, it doesn't get to the core of the problem, which is precisely that dependency. A corporation whose existence is the pre-condition for the existence of others is defacto a monopoly. It is considered in the US as "too big to fail", but shouldn't that be an argument to the contrary: "too big to stay as it is".

It's not only small businesses that have become dependent of Big Tech, but the government as well. Amazon Web Services provides the cloud servers that powers the government data, Google provides the search technology that provides access to government data, Apple and Google combined provide the totality of cell phone traffic, including the government information.

As long as the government is not able to operate without depending on these companies, its leverage will be extremely limited. Anti-monopoly measures will be window-dressing. Congress may successfully impose Google and Apple to provide equal access to the various apps on the marketplace, but that won't provide a dent to their power. As their algorithms are not accountable, they still have plenty of leverage on the page ranking or app ranking to be able to comply with Congress requirements for inclusion while continuing to subtilely privileging their chosen apps. But they are able to use the government action to mobilize their users in defiance to the government.

The core business model, based on reselling their users' data, is not touched by these anti-monopoly bills. Citizen privacy will remain unreachable, and in addition citizens will increasingly lose confidence in the ability of the government to act in their favor.

Europe's efforts are too shy

In Europe, recent history has made citizens more sensitive to surveillance, both by governments and by corporations. Privacy is considered a condition for freedom, more than in the United States, and the GDPR has been implemented to protect citizens from abuse coming from Big Tech. But the problem is that most of the technology used in Europe is not produced in Europe. European countries do add tax burdens and privacy regulations to US-based companies. Because the Internet is a world-wide operation, this attitude has an impact on the US market as well. But it doesn't modify deeply the situation, and the fines that US companies have to pay are not preventing them to continue operating the way they choose.

Silicon Valley overfishes

Research and innovation, which flourished during the 1990s especially in the Silicon Valley, with the emergence of the Web, e-Commerce, social media platforms, and office software applications, has been so successful that massive consolidation has occurred. A handful of companies have been swallowing smaller, innovative companies and have reached a size and an economic power that is unprecedented. At the same time, they have absorbed most of the intellectual resources. Brilliant researchers, industry visionaries, intellectuals have been hired in the research departments of these companies, that allow them to develop more products and continue to grow. There are two downsides to that evolution: first, the research activities conducted by them are not contradicting their core business model. On the contrary, they are reinforcing it. Second, they have dried up the competition. It has become increasingly difficult to conduct research project outside of these giant corporations, because they own the resources on which every potential competitor depends. Amazon for example has found out that ElasticSearch, which was hosted on their server, did something that was of great interest to them. They stole the code, and proposed a deal that translated into a huge loss for ElasticSearch. By sucking up their potential competitors, they have dried up the field. The investors have followed them, and are only interested by funding companies that show a potential to become as big as the happy few, therefore preventing any smaller companies to thrive. As the Big Tech companies have also had a disruptive effect on the real estate in the Valley and surrounding areas, they have caused an exodus out from Northern California to other regions. It is not clear that the Silicon Valley will

be anything else than the historic headquarters of companies that flourished at the beginning of the twenty-first century. If the Valley is not able to reinvent itself, it may just stop being a reference point for innovation.

Attacks against democracy

The first amendment protection has been perversely used by big tech corporations. Freedom is good as long as it lets them whatever they please, without any accountability mechanism. It doesn't matter if there are negative consequences, such as the inability of distinguishing true from false, right from wrong, fair from unfair, peaceful from violent, fact from fantasy. Anything that can be used to limit the ability of social media platforms to develop their business by attracting people to stay on their platform, be it with valuable content, or outrageous content, is being fought tooth and nail. Marc Zuckerberg is the best representative of this trend.

The US Government, traditionally, has a very broad notion of preserving freedoms. The freedom to bear arms, for example, is something sacred, even if it results in thousands of deaths and terror scares in places that should be safe heavens, such as schools, churches, shopping malls, etc. Hate speech is protected by the first amendment. The limitations to those freedoms is reached - sometimes - when violent acts actually occur. But not always. It then becomes an influence war between the gun lobby and the anti-gun activists, between the big tech giants and those who try to protect the rights of the victims of bullying, harassment, prevent suicides, or defend the truth between lies, etc.

As Shoshana Zuboff has described, the protection of the laissez-faire in tech doesn't depend on the political party. An alliance occurred between the Republican Administration after 9/11 to get as much information on citizens as possible from Google. In exchange the government accepts to renounce to any kind of regulation that would prevent Google to exert their practice of surveillance in full freedom. In 2012, the Obama campaign made a similar deal with Facebook. In exchange for getting information about the voting preferences of the electorate, the democratic administration committed in not interfering with the ability for Facebook to gather as much information from their users as they wanted. [^zuboff]

[^zuboff]: Shoshana Zuboff, *The Age of Surveillance Capitalism*.

Nobody is there to protect the rights of citizens-users to keep their personal information private. Zuckerberg and others say: Privacy is dead, as it is was a fact that was made inevitable by technology. It is only there because nothing prevents him, and others, to do as they please, regardless of the negative consequences it may have on millions of people. As Joshua Fairfied puts it, the inevitability of technology is purely ideological.[^runaway-tech] One doesn't have to be a Luddite to claim that it should be possible to benefit from technology and do it in a comfortable way, that preserves our individual privacy, and eventually, freedom to do what we want, read what we want, and think the way we want, without being reported to big tech corporations, the government, or both.

[^runaway-tech]: Joshua Fairfield, *Runaway Technology: Can Law Keep up*, 2021.

As long as the surveillance will keep going on unhinged, it will favor places where government can use surveillance tools designed to capture individual information. The authoritarian regimes have the ability to appear up-to-date, and therefore attract a trove of young people, who are into technology, because they favor the use of technology. China, for example, uses a "social credit score" to allow or prevent citizens to access certain essential services that are considered obvious rights elsewhere. For example, Chinese citizens who have a low social credit rating, if they don't actively give enough "likes" to their government on social media, are prevented from the right to travel [^wichowski]

[^wichowski]: Alexis Wichowski, *The Information Trade*, Harper & Collins.

2. Democracy rebounces

- increase awareness of the risks
- tech development considered a priority
- research on how to maintain a democratic society fueled by technology
- new accountability features adapted to tech-driven society.
- citizens have freedom of speech, increased privacy rights, adjusted property rights

- accessible technology considered a national priority.
- dismantling of excess of power gained by tech giants
- creating an ecosystem in which competition and innovation are encouraged at all levels of the society (not just for minority-owned, etc. businesses)
- universality
- increased education level: right to education recognized as the right to be healthy.

Focus on data management.

- Big Data = Change of paradigm (see J. Fairfield)
- Content quality vs. AI and ML
- Relation with liability on content and the disruption of the traditional publishing model by social media platforms.
- XML vs. JSON
- RDF vs. Topic Maps
- Standards vs. de facto standards. Word/PDF vs. XML.
- The browser war.

Conclusion:

- Lots of work needed. Need for creating a NIT (equivalent of NIH for technology).

It's not looking good.

Freedom of speech is being abused and enables hate speech, bullying, harassment, blackmail, conspiracy theories, and Russia-backed propaganda to disseminate all over the country. Citizen privacy is not guaranteed, individual data is being used, sold to advertisers. Critical infrastructure and IOT-enabled devices are subject to hacking. Most people are feeling powerless in face of what is described as the inevitable consequences of technological evolution.

It's big, and looks unstoppable. But it's not unstoppable. It's simply not being stopped. However, authoritarian regimes are able to muzzle technology, by tweaking search results to hide for example dissidents' ability to publish information

on the Web. Europe is actively implementing solutions to protect their citizens (e.g. GDPR). Nothing similar is happening in the US.

US is losing its lead on technology world-wide. US is losing control over how technology is being deployed domestically. US companies are subject to anti-regulatory actions abroad (last is India vs. Apple [^Newley-Purnell-2022-01-03]).

[^Newley-Purnell-2022-01-03]: Newley Purnell, "India Hits Apple With Antitrust Investigation Over App-Store Practices", *Wall Street Journal*, January 3, 2022. https://www.wsj.com/articles/india-hits-apple-with-antitrust-investigation-over-app-store-practices-11641207296?mod=hp_lead_pos4

More specifically,

- cloud services which provide critical infrastructure are entirely under the control of the private sector: Amazon's AWS, Microsoft's Azure, Google's Cloud, are the three leaders in Cloud services.
- search engines: In 2021, Google occupies 91.42% of the market of search, Microsoft's Bing occupies the second place with only 3% of the search. [^search-engines]

[^search-engines]: According to: https://www.reliablesoft.net/top-10-search-engines-in-the-world/

The Decline of Silicon Valley

The case of the NIH

The US carries out 46% of global research and development (R&D) in the life sciences, making it the world leader in medical research.[^USA-life-sciences]

[^USA-life-sciences]:Shannon; Springs, Stacy (2015). United States of America. In: UNESCO Science Report: towards 2030 (PDF). Paris: UNESCO. ISBN 978-92-3-100129-1. Quoted in Wikipedia article: https://en.wikipedia.org/wiki/Biomedical_research_in_the_United_States

US needs government-funded technology

- Why? Because US is competing against other countries that develop their own national technological infrastructure. China.
- How? Two existing institutions:
  - NIH
  - Post Office (mail services)

Who can benefit ?

- The public:
  - Reliable, privacy, freedom (e.g. USPS mail vs. Google Gmail)

Who is against?

- Unregulated Big Tech giants.

It already exists: Case of NIH and Post Office (mail)

It doesn't exist.

No public infrastructure

Cloud and server farms

Access / Search

Communications: Social media platforms, Email services.

No equivalent in technology

The Obstacles

Research is mostly happening within big tech companies using their own inf

Screaming / Dreaming

Americans Dreaming

Caffee Caffee

Old-fashioned tearoom. Wooden walls.

- Adam: The Roman Empire vanished. America is disappearing. So what?
- Me: How can you be so cynical? Is there nothing to do to revert the fall? To prevent disaster from happening?

Chapter 1. Screaming and Dreaming

There are plenty of opportunities to scream. Let's do it. Let's scream, as loudly as possible. Let's not try to be polite, to be civil. Or more exactly, let's not leave that right exclusively to the other camp. Enough is enough.

The Big Undoings

Destruction and violence make big success stories for shows. Big shows use panic, awe, anxiety. We have reached a point where those who conceive the shows have checked all the boxes to be extremely successful. But what if what was initially conceived as a technique for getting the audience riveted becomes real? What is there a point of non-return where we suddenly realize that it's not a show any more, it's the reality we live in. Is it already too late to reverse course?

Democracy

Let's imagine that democracy is too complicated to be handled properly. There are too many stakeholders, with conflicting interests. The negotiations, compromises, discussions are too hard to handle. Democracy is a swamp. Let's eliminate it. [... SCREAMING]

Let's imagine that our society is too complex to grasp. We can simplify it and find simpler, straightforward, obvious answers. Imagine that instead of allowing multiple opinions that result in constant bickering, we recognize that it is more efficient when there is one thought, one leader, one people of the same color, the same religion, the same ethnicity. Let's eliminate anything that doesn't fit. Let's regroup under our dear leader and let him do the right thing. We could get a break if instead of thinking, we just follow orders. [... SCREAMING]

Let's imagine that checks and balances are there just to prevent our dear leader to do what he wants. Congress doesn't accomplish much, and the judiciary system is rigged. We could get rid of them, and get one executive be the head of the executive branch. That seems logical, reasonable, and it should work much better. [... SCREAMING]

Let's imagine that elections are useless. It's a huge waste of time and money. Moreover, people from one party get elected and can't even decide to do what their constituents want, so why bother? We could terminate that party, stop being

distracted and save our time and resources to work efficiently under a stable and permanent leadership, without constantly criticizing it. [... SCREAMING]

Constitution

Let's imagine that we keep the first, second and fifth amendments and ignore the others.[... SCREAMING]

Let's image that we don't need to talk to people we disagree with, or who are not like us. If they are not like us, it means they don't like us. We don't have to like them. We can hate them. We have the right to do so, it's protected by the first amendment.[... SCREAMING]

Let's imagine that the second amendment should be applied more broadly. Everybody has the right to bear arms, regardless of whether it is to get rid of enemies, regardless of whether we are in good mental health or not, or whether we just feel like killing others is a good idea.[... SCREAMING]

Let's imagine that we don't need justice to force us to show up and to talk if we don't want to. Let's ignore the law and plead the fifth anytime we want.[... SCREAMING]

Science and Technology

Let's imagine that the forests are burning, coastal areas get regularly flooded, hurricanes are more frequent and violent. There is nothing we can do about it. Let's not waste any more resources in climate science.[... SCREAMING]

Let's imagine that we refuse any government interference with our bodies. For example, getting vaccinated would be a free choice. However, there are cases where we need the government to tell us what to do: Let's imagine that only the government should tell women that they are not to that they should do with their bodies, as they are the ones who can decide if they want children or not. Let's not waste any more resource in medical science.[... SCREAMING]

Let's imagine that computers are more efficient than humans. Let's have robots, drones, let's get AI-driven algorithms manage our affairs for us. What we need to do

is to learn to think like computers: everything has to be true or false, nothing else.[... SCREAMING]

Let's imagine that privacy is dead. We should get over it.[... SCREAMING]

The undoings

- Undoing the 1960s: Why make love if we can make war? Why make civil rights if apartheid is more viable? Why should women not choose to die if they can't get through pregnancy or to carry the child after having been raped?[... SCREAMING]
- Undoing the 1950s: Why maintain a world order based on the United States being the protection against authoritarian regimes? Why pour money and resources into other countries? What is there for us, in exchange? [... SCREAMING]
- Undoing the 1940s: Why combat against fascist regimes when they provide answers that could work here? [... SCREAMING]
- Undoing the 1930s: Why keep entitlements when they encourage people to be lazy and wait for their government checks instead of going back to work? [... SCREAMING]
- Redoing the 1920s: Why protect us from pandemics when they are a natural way to eliminate overpopulation? Why regulate anything when the market invisible hand is the miracle solution that always works, no matter what. What about make lynching legal again? [... SCREAMING]
- Redoing the 1910s: Why not split the countries into smaller, more homogeneous units, that will be able to go to war against their neighbors, for any reason they choose. The difference, though, is that we could start playing with our nuclear toys.[... SCREAMING]

Fighting

Done with screaming. Suppose that everything that precedes is not the fruit of a sick, paranoid, dystopian, feeble character, but this is — unfortunately — a

reflection of the issues we are facing and that we need to address.

It is easy to be overwhelmed, lost, and voiceless when thinking of the enormity of the task, especially if we believe that most of these issues are inevitable. But at least we can concentrate of one of the alleged reasons why this is happening. We often hear that technology is what makes this inevitable.

Nothing is further from the truth. Recent works are showing that what we think as inevitable, for example the death of privacy, or the propagation of hate speech, are not consequences of technology, but they are arguments in defense of special interests, that want us to believe this to continue maximizing their profits and ensure their domination.

I will mention three books.

Shoshana Zuboff, in *The Age of Surveillance Capitalism* brilliantly exposes how a new form of economy has been created from scratch, due to accidental circumstances. In the dawn of the twenty-first century, Google was looking for a viable business model for their Web search engine. They discover that by collecting the "garbage" left by their users on computers, they could gain a very intimate knowledge of what they were thinking, not only what they were searching, and they had a gold mine that could be sold to the advertisers looking for very precise market segments. 9/11 happened and the government made a deal: they wanted to get that information from Google. In exchange, they promised they wouldn't do anything to regulate the fact that Google was getting a lot of information from their users, without telling them.

Alexis Wichowski, in *The Information Trade*, shows how the Big Tech companies created what she called "net states", entities that can rival and sometimes surpass the power of "nation states".

Joshua Fairfield, in *Runaway Tech: Can Law Keep Up* shows that it is absurd to claim that law is not able to keep up with technology, because law itself is a technology and proposes ways to adapt the law to be able to act on the challenges brought by technology.

In what follows, I will show that nothing that has been said so far is inevitable and that it is possible to take a big breath and conceive concrete and workable solutions.

The first step is to recognize that this is a fight against people, companies and organizations which have vested interests to continue the current state of affairs. The fight is not against one unique entity. There are many important differences, and conflicts between the protagonists.

Here are some of them.

Authoritarian regimes have a propensity to control everything. Technology is not only regulated, it is entirely under the command of several governments. In China, for example, Alexis Wichowski describes the system of "social credit score", which is more or less a mandatory system of "likes" that Chinese citizens are supposed to give in support of government initiatives. The social credit score is a numerical index, and it opens the way to "rewards", which are in other places basic rights. For example, people who don't have a credit scope above a certain threshold do not have the right to travel.

Russia is able to control search engines to exclude a number of web sites from being found, including those created by dissidents.

At the same time, in the United States, the Big Tech companies are totally unregulated, and the United States government, together with the Big Tech companies, make people believe that technology prevents any filtering from occurring. This is obviously not the case, since the authoritarian regimes show that when they want something, they can get it done. The truth here is that the US Government doesn't want to do anything because it is their interest to preserve the status quo with the Big Tech companies.

New York City just announced that they are auditing artificial intelligence algorithms. Anything that can not be audited will not be allowed to be used in relation with official government business. This is a gesture that does not signal a tendency towards authoritarianism, contrarily to what some people may say. It indicates a willingness to keep technology under scrutiny because the public has the right to

know. This seems to be exactly the kind of accountability that is the condition for a democratic society to occur.

The book publishing industry, which exists since more than five centuries, has flourished in free societies. However, a publisher is exposed to liabilities in certain cases, like libel, or privacy violation. A person that considers that its reputation has been damaged by a publication has the ability to sue a publisher and legal provisions exist to discuss each issue. The problem with a social media platform such as Facebook is that Facebook opens its platform to anyone who wants to publish something, with a blurry concept of what is acceptable, and no equivalent liability provision. Facebook itself declines to be considered responsible for the content, and the authors of the content are hard to find, and are using the first amendment as a general protection, making law suits difficult. This is a domain worth exploring in more detail, and laws should be designed, when and if necessary, to address some of the thorny issues.

The right to vote is a contentious issue these days, for cynical reasons, as one party considers that if they play the game as it is, their chances of winning are meager. They have spent a great deal of energy trying to convince people that the electoral system is rigged, that the voting machines can be tweaked, etc. Although this may be far from the truth, they have successfully cast a doubt on the electoral process. Here, one possible approach could be to impose that software used in voting machines be open-sourced, so that it could be scrutinized at any desirable level of detail by tech savvy citizens, who, as a community, could discuss whether or not there are flaws in the technology.

Another area for scrutiny is the way the Big Data paradigm is being implemented. It is described as a magic black box. Consumers of artificial intelligence and deep learning algorithms are asked to blindly trust the software, often without the ability to look inside it. This also applies to search. The page ranking algorithm used by Google is entirely secret. Google makes its own rules, and presents its constraints under the expression "search engine optimization". For a company, complying with Google's constraints may be as complex, or maybe even more, than filing a tax return. The difference is that the tax return is based on laws that can be read, analyzed, and even opposed to, in case of a situation where a taxpayer disagrees

with the IRS. There is no case where a Google search engine user can "disagree with Google" even if the consequences of a page ranking change may be fatal for a company that is no more found, for unknown reasons. The remedy here is not to impose Google to publish the internals of their search engine — they will never accept to do it —, but to create a competing public search engine that can be used instead of Google and would be designed in a way that could be open to discussions with the users.

When we send a letter by snail mail, the Post office is not opening our letter, to check what we write so that they could send us advertisement that correspond to the content of our letter. In fact, privacy of personal correspondence is considered a must. It's only in case a person is under legal investigation that an order can be put in place to seize their mail. Nevertheless, the email we send using big tech companies - Google, again, is the most used email system - we know that Google reads our mail. It's nothing person, just for advertisement purpose. But we have no idea to whom they sale the email content. Even if it's anonymized, they keep also track of the metadata (who sent an email to whom). It's only in fascist or communist dictatorships that the mail was open. And this is what we are facing everyday when we use Gmail. Again, there is nothing inevitable here. It is not unreasonable to imagine that there be a government-supported email service, exactly like the mail service, that could be even be handled by the USPS, that offers a high quality service with the guarantee that under normal circumstances, the correspondence remains private and is only open under a well maintained list of specific situations. Again, a citizen who considers that their privacy is violated because an email has been opened when it should not have been, would have a recourse.

Dreaming

After screaming, it's time for dreaming.

- What if technology would be something that could benefit humankind?
- What if we create the intellectual weapons to resist against hopelessness and inevitability?

- What if we would agree to disagree, be comfortable with it, and call it a democracy?
- What if we question the binary model used by computers and point its limitations?
- What if we get to recognize that artificial intelligence has nothing to do with human intelligence, and even if some jobs could be mechanized, the imagination coming from human intelligence is limitless?
- What if we would reinvent democracy to incorporate novelty, technology?
- What if we work to make people's lives better instead of worse?
- What if human life is the most precious thing we care about?
- What if we work on enjoying the planet we live in and make sure it stays livable?
- What if America would be a place where science is being encouraged, serves the public good and regains a leading place in the world?
- What if we use technology to work on getting closer to the fact that men - and women - are created equal?

Initiatives are encouraging. Coming from multiple parts of the country, including New York City government, the Center for Humane Technology, and many publications around those subjects.

Bold dream

The US has a National Institute of Health, which the largest biomedical research agency in the world, "making important discoveries that improve health and save lives".

Research in technology is conducted inside big tech corporations and in a number of universities. There is no national research agency on technology equivalent to the NIH.

We advocate for the creation of the National Institute of Technology. Products from NIT, such as a search engine, email service, cloud storage service, would be at

least on par, or on the leading edge, compared to similar research in big tech companies. Government agencies, and the public, could benefit directly from the services provided by NIT. Congress members could have a place to go in order to learn and understand in much granularity the challenges of today's and tomorrow's technologies.

Knowledge Graph Catch 22

Michel Biezunski, Infoloom

January 15, 2022

The raison d'être of knowledge graphs is to provide the ability to connect any data point to any other, without being constrained by an existing data model.

However, this freedom creates loose environments, and any application that is useful in a given work environment needs to be built from scratch each time. Although it is technically possible using graph databases, the amount of work needed to get to a useful technological environment is quite high.

The triple approach provided by the Semantic Web applications built on top of RDF (The Resource Description Framework) helps ease that burden. Every link in the graph follows a pattern designed as "subject - object - predicate". This kind of knowledge graph has the ability to be used to perform automatic calculations, created by designing inferences using the triples as the raw material. These inferences use the "sentences" expressed with triples.

Another viable approach used within knowledge graph consists in designing compound objects for the nodes. For example, a person node is defined with explicitly declared properties such as name, address, email, etc. An address is itself a compound object, having a street, city, zip code, etc.

Each of these approaches have drawbacks. The triple approach is based on the premise that every link can be expressed with a sentence conforming to the equivalent of a simple grammar rule, such as subject, verb, complement. Not all the relationships follow that pattern, and it is sometimes a stretch to express things that way.

The compound object approach defeats the flexibility expectation, as nodes have to conform to micro-structures which are strictly predefined. The compliance is deferred to multiple micro-levels.

The solution is to look for an architecture that both open the ability to connect anything to anyhting else, but that still has a generic structure that is useful and makes sense maybe not in all cases, but in a significant high number of cases.

A node can be used to describe a topic in a generic way, that expresses anything we can talk about. We can give it a name, or even several names. The name itself can have a special attribute used for disambiguation, as in New York (State) versus New York (City) versus New York (County). It can additionally be provided with one or several language values that allows the node names to be filtered according to languages. For example, New York (English, French, Italian, etc.) vs. Nueva York (Spanish).

A node can be related to other nodes using relationships whose semantics can be characterized by another node equipped with the same properties, which gives the possibility to disambiguate the name used to qualify the relationships, as well as the ability to express the semantic in other languages. This way to express a knowledge graph is based on concepts that were originally developed in an architecture called Topic Maps.

This view of a knowledge graph still enables full flexibility in defining the information set. It levels the difference between data and metadata, by considering that a metadata attached to some data is another kind of link. It also opens the ability to run powerful processes in the graph, by issuing queries based for example on neighbors to express specific semantics.

The internal structure of complex objects becomes accessible as post-processing. But it is not as rigid. For example, an address may contain a "field" (i.e. a relation) with the semantic "province" instead of "state" if it's more appropriate in a given context, even if that "field" didn't pre-exist.

Focus

Small Data

As data extraction is considered the new golden egg, big data technologies are running the show. Algorithms are used to extract data, and data science is used to get inside the data and target the data points that are relevant. For example, getting to know what potential customers want, how they feel, what they buy, and how much they pay for it is a powerful tool for companies aiming at selling their products. Therefore, the advertising industry relies on data accumulated by social media platforms, cloud infrastructure platforms, email, all kinds of communications.

Exploitation of data is the basis for a new kind of economic activity, described by Shoshana Zuboff as surveillance capitalism, which explains the process by which a handful tech companies have been able to reach a level of wealth and influence that is unprecedented. Tracking technology is also used to reinforce the power of authoritarian regimes, that embrace technology and run the show. Wichowski is showing the complex relationship between the nation-states and what she called net-states to describe the power that has been accumulated by big tech.

Artificial intelligence and deep learning algorithms are considered today at the center of any technological infrastructure. There is too much data available to handle it in any other way, and these techniques also open the door to the so-called Internet of Things, or the Smart Economy, where many devices are now connected and extract and produce data: smart phones, watches, fridges, TV sets, cars.

Critical infrastructure, such as the power grid, the government, the financial system, healthcare, etc. all are built or adjusted to use these techniques.

Something big is missing. Humans are the ones who consume data and base their decision process on the knowledge they have. Knowledge includes access to data, but also the ability to figure out what to do with data.

Sustainability, community-driven activities, coopetition, carbon print awareness, etc.

Can technology be tamed?

Where are we and where are we going.

The main question in this book is to debunk the idea that technology runs the show and inevitably dictates any decision made by an individual, a corporation or a nation.

Instead, we will use a specific aspect of technology —how data is managed— to show how an ideology is being constructed by those who are looking for reinforcing corporate power, weakening democracies, and eventually demoralizing individuals by convincing them that they have lost any power to do anything about it.

Therefore, the question is not so much to educate people who would evolve if they would know more about how technology works. It is rather aimed at fighting against those who allegedly know what they are doing. It is important not to underestimate the enemy, but instead to outsmart them by showing that technology also offer solutions that can be used to resist against the main trends. This is a discussion which parallels the right to vote. There are people whose interest is to prevent

others to vote. The first thing to realize is that this is what is happening, not to lament that they don't understand the rules. This lesson has been learned painfully from the Trump presidency. At the beginning, "reasonable" people were shocked by his behavior, and thought that his inexperience led him to favor unusual decisions, such as supporting Putin and other dictators. Unfortunately the truth was that he knew what he was doing, and behaved in a consistent way, profiting from the fact that nobody was fighting against him because they didn't consider him seriously. The situation is similar with technology. For example, the fact that Google and Facebook can profit from selling private data to the biggest bidder is not a normal consequence of technology. They can do it, but, more importantly, nobody prevents them to do it. And their interest is to make everybody believe that there is nothing they can do about it, because this is the way technology goes. It is not. Or rather, they have created a powerful, and vastly shared, ideology that helps them assert it.

At the individual level, the idea that one person's integrity and privacy can still be maintained seems to fade out of the possibilities. At the corporation level, the idea that AI and deep learning are the only ways to manage data leaves out many use cases, especially when a company creates their own products and data and is willing to retain control over how to manage well-known data points. Furthermore, the idea that freedom of speech prevents any interference from the government to regulate content provided by technology company jeopardizes the normal operation of democracy. On the other end, the idea that technology dictates the rules reinforces authoritarian trends especially when a nation-state decides to control technology.

The context

This book builds from previous works that represent important intellectual breakthroughs:

Shoshana Zuboff brilliantly shows that the business model that has made Google and Facebook so powerful is not based on a natural technological evolution, but is the result of accidental opportunities. These companies have been smart enough to exploit the opportunities to their limits, and ensuring by various means that they

would not accept to be stopped implementing them. Google has discovered that there was value in the garbage left on computers by users that enabled them to collect intimate knowledge on people, and that the trove they got had a marketable value because that data could be sold for advertising purposes. The US government was interested by using data acquired by Google after 9/11 to track terrorist activities and made a deal with Google to that effect.

Facebook, similarly, discovered that people were sharing lots of intimate details that they could exploit for advertising purposes as well. The ability to publish online and reach out various groups of people based on common interests was of interest for the Obama campaign in 2012 and they renewed the deal to acquire data while ensuring that Facebook could continue collecting whatever data they wanted, without asking permission, or making it in a way that was convoluted enough to discourage people to properly set them.

Shoshana Zuboff [^zuboff] shows that Google and Facebook business models can only thrive as long as they are unregulated and that the efforts of those two tech giants is focused on ensuring that no significant dent is made against their capacity to collect and resell consumers' data.

[^zuboff]: Shoshana Zuboff, *The Age of Surveillance Capitalism*.

Alexis Wichowski [^wichowski] shows that the nation-states are competing with big tech companies that she calls "net-state". The authoritarian regimes are able to reign it tech companies and use technology in an innovative way to control their citizens. For example, the "social credit score" invented in China is similar to the credit score used by financial institutions in the United States and is used to measure the degree of compliance of citizens towards the rulers.

Alexis Wichowski shows that dictatorships and authoritarian regimes are able to reign in technological companies, by imposing them to implement rules that serve their interests. Although the same companies are making people in the US believe that it is too complicated to filter some information out (for example, tweak the Google Search algorithms), they have much less problems to comply with the requirements set up by authoritarian regimes, who threaten them to block their access if they don't abide by their rules.

[^wichowski]: Alexis Wichowski, *The Information Trade*

Joshua Fairfield [^fairfield] shows that there is a way out to the conundrum that law can't keep up with technology. He shows that law itself is a technology, and shows the paradigm shift that occurs between algorithms producing data and data analysis that creates conclusion directly from collected data.

[^fairfield]: Joshua Fairfield, *Runaway Tech — Can Law Keep Up*

In order to illustrate the main point of this book, I will start with the thought that what is happening is not the natural result of the evolution of technology, but results from a conscious willingness to favor certain trends, and uses technology to that goal. I will show that this

Why authoritarianism is so attractive

Simple to grasp

Whether the issues are simple or complex, the answers are simple. They can be summarized in easy-to-remember zingers.

Efficient and cost effective

It's the natural way to go to get efficient tech. One model, accepted by all. Universal "agreement". Ambiguity not allowed. This facilitates computer processing. Instead of spending countless hours compromising by integrating all kinds of points of view, let's have everybody adopt one vision. Computers are more efficient in dealing with it.

Spare the masses from exposure to the internals

Technology is inherently complex. It's better handled within black boxes. The less people know how it works internally, the easiest it is to maintain and evolve.

Ability to monitor users

Users are tracked in all their gestures. Every click can be recorded.

Reach out the masses

Social media are honey traps that let people tell everything about themselves. They think they share their data, feelings, opinions, with their close knit communities. In reality, they transmit information to those in charge of the systems.

Two ideal contexts

State authoritarianism: Ex. China Unregulated Corporations: Ex. USA

In China, the state is in charge of controlling the population, and tracks all data they can in order to maintain power. Citizens are not able to express any disagreement.

In the USA, the big tech corporations collect data on their users that they can monetize for commercial purposes. Corporations operate in a mainly unregulated environment where the users have lost all power.

Why democracies are failing

Complex problems.

Problems affecting mankind and its future are not simple problems that have simple solutions.

Science does not have all answers. There are controversies, and disagreements, within the scientific community. Therefore, there is no unique way to distinguish right from wrong. It takes intellectual courage and creativity to admit that there is an endless quest with no certainty.

Managing divergent opinions is long, costly, and unsexy

Endless meetings, negotiations, discussions where divergent perspectives are exposed can be tedious and frustrating. At the end, compromises can generate

frustrating from all sides. There is no point where things get settled for good. There is no end to this process. Things are never completely black and white, but there is an infinite variation of colors involved.

Spare the masses

Unequal access to education and power creates big differences. Decision makers are a minority and their decisions affect the masses. They have, theoretically, the right to propose alternatives if they don't like what they see, but they may not have all the information available to oppose to the rulers.

The carrot and the stick

The "recommendation" system is the clearest example of why people fall for getting their data extracted from them. If computer systems are able to detect what we like, it will show us other things we need, and that make us feel good. Surveillance is therefore appreciated by their victims.

Social media and information sources

As everybody not only can publish but have their information published on social media look "professional", people believe what they see written and published. Fake news, disinformation, hate speeches, etc., are practically indistinguishable from vetted information. In addition, the fact that the liability encurred by traditional publishers has no online equivalent, it's much harder to prevent lies and hate speeches to propagate.

Is technology responsible

Technology has two sides: positive and negative. On the one hand, it provides unprecedented access to tons of information that improve communication between people, across borders and reduce distance, allow for sharing video calls that put people together, simplifies many procedures that make life much easier. On the other hand, technology also provides easy access to political propaganda,

increases hate speeches, bullying, remove privacy, makes a lot of systems more vulnerable and prone to attacks.

Surveillance is a historic accident.

- See Zuboff about Google, the US Government and 9/11.

An authoritarian ideology based on technology

Preaching that the way things are set is a direct consequence of technology.

Debunking: this is reinforced by big tech companies which want to preserve their business models

Solutions

- Perfect Order: Mess is also part of life. Ambiguities, double meanings, various interpretations, critical view points, multiple perspectives exist. Even if it doesn't work well for computers, they are essential for humans.
- Big Data. Big Data is not all what there is. Small Data also exists.
- Artificial Intelligence. Paradigm shift. Algorithms are of two kinds:
    i. logical reasoning, producing data from hypotheses.
    ii. collecting data and building dashboards from data. Not answering the questions of why.
- Deep Learning. Machines getting out of control. Frankenstein effect.
- Terms of use and Consumer Protection.
- Rights to privacy, right to be forgotten

World order

- Democracies are failing: technically and ideologically. USA diminished role vs. China. The reason: lack of understanding what technology is about, due to the willingness of big tech to stay unregulated. Parallel to other flaws in the American society: belief that healthcare is a product that is being purchased according to each person's wealth, belief that having a gun is a freedom that needs to be protected. Third: right for big tech corps to do as they please, to preserve free market from regulations.
- Rather than imposing anti-trust laws that often are just fig leafs, US would be better off if the political and conceptual understanding of how technology works would be significantly improved. Technology can be developed with the public interest in mind. For example, it should be possible to deploy a national email service (even handled by USPS) that works like gmail, but where the correspondence between people would be guaranteed to stay private (unless in specific cases defined by law, such as when a person is under investigation, etc.)

Radical solution

- For open and democratic societies to survive, they have to recognize their current weaknesses and what they are against. In authoritarian societies, people are excluded from any decision making. They are not allowed to think. People have no voice in societies where corporations are unregulated. By people, I mean its users as well as its workers.
- Technology, however complex it has grown to become, can't be left without control by citizens of a democratic society. We need to understand how it operates, how it is deployed, what are the reasons for it to be the way it is, and to express an opinion on possible alternatives. For example, it is possible to communicate via email, text, or video call without being tracked. For example, it should be possible to use social media sharing information limited to a group of people we choose. It should be possible to publish people using publishers' platforms that validate, vet, and are taking full responsibility for the content they publish. It should be possible to imagine that a country voting system uses open source software that is open for review by all techno-citizens who

collectively guarantees its integrity. It should be possible to use cloud platforms without given away any property rights to the hosting company. Nothing forces us to consider that there is something inevitable about the way things are now.
- Societies which invest in public interest technology are rewarded by being considered to be technology leaders in the world. Especially if they offer protections that ensure that the public can benefit from the advantages of technology without having to be taken hostage by a handful of giant corporations which thrive precisely as long as they are not regulated. The US is recognized in the world thanks in part to the existence of its National Institute of Health which uses federal funds to conduct research that benefit, indirectly only, to the corporations in the sector. Research and development is in its majority

Focus on publishing.

Role of publishing in law

- Common law relies on *published* cases.
- Civil law relies on *published* statutes.

Common law and civil law systems predate the printing press. What has printing technologies changed in the ways law is applied and the rule of law in general. In which way modern democracies (i.e. created after the printing press) are impacted by the use of printing. What roles publishers play?

Laws about publishing

Alasdair Taylor, *10 things you should know about … publishing law*, 23 Jul 2012, 10 https://seqlegal.com/blog/10-things-you-should-know-about-publishing-law

8. Forms of content liability

There are many different ways that legal rights can be infringed, and many different sorts of legal wrongs that can be committed, by the simple act of publishing a written work. For example, a single work could: be libellous or maliciously false; be obscene or indecent; infringe copyright, moral rights, database rights, trade mark rights, design rights, rights in passing off, or other intellectual property rights; infringe rights of confidence, rights of privacy, or rights under data protection legislation; constitute negligent advice; constitute an incitement to commit a crime; be in contempt of court, or in breach of a court order; be in breach of racial or religious hatred or discrimination legislation; be blasphemous; or be in breach of official secrets legislation.

9. Moral rights

Moral rights arise in relation to most works that attract the protection of copyright. Unlike copyright moral rights cannot ordinarily be transferred, although as a matter of English law at least they can be waived. The most important moral rights are the right of paternity (i.e. attribution), the right to object to the derogatory treatment of a work, and the right to object to the false attribution of a work.

https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199987443.001.0001/acprof-9780199987443-chapter-4 The Printing Press and Constitutional Revolutions Maxwell A. Cameron

Strong Constitutions Social-Cognitive Origins of the Separation of Powers Strong Constitutions: Social-Cognitive Origins of the Separation of Powers Maxwell Cameron

Print publication date: 2013 Print ISBN-13: 9780199987443 Published to Oxford Scholarship Online: September 2013 DOI:

10.1093/acprof:oso/9780199987443.001.0001
DOI:10.1093/acprof:oso/9780199987443.003.0004

The printing press contributed to the spread of reading and writing. Although the initial impact of the Gutenberg revolution was to reinforce monopolies of knowledge, over time it contributed to the separation of church and state, the emergence of public opinion, and the differentiation of powers within the state, including the rise of parliaments as legislative – legal text-making – bodies. Theorizing about the role of legislatures, courts, and executives was sharpened by Enlightenment thinkers – among them Montesquieu, who provided the canonical synthesis. The new contractualism opened the door to constitution-making as a conscious, adaptive, and evolutionary activity. This new spirit informed both the French and American revolutions, as well as constitutional traditions in those parts of the world colonized by the West. In particular, they provided the initial point of departure for both the Madisonian tradition in North America and the Bolivarian tradition of Latin America.

Press vs. Democracy

Michael Luo, "How can the Press Best Serve a Democratic Society", *New Yorker*, July 11, 2020. (saved)

Elisabeth Eisenstein, *The Printing Press as an Agent of Change*

Effects of Print Media on Society, https://www.zeepedia.com/read.php?effects_of_print_media_on_society_economic_ideas_politics_introduction_to_mass_communication&b=78&c=19

Printing Press- The Internet of middle ages Brief Summary of Impact of Major Information Technologies Bikesh Singh
https://dumbcontract.substack.com/p/impact-of-printing-press

Opinion: Mark Zuckerberg, meet Johannes Gutenberg, https://www.washingtonpost.com/opinions/2021/11/09/zuckerberg-gutenberg-printing-press-social-media-disruption/

Nadia Eghbal, Working in Public: The Making and Maintenance of Open Source Software, Stripe Press, 2020

The adventures of Tom Picmap

Imaginary biography of a character that went through the period of recent upheaval in tech.

Using Topic Maps as a thermometer, or an angle for other things.

- on China competition with US
- on accountability
- on freedom of speech
- on authoritarianism
- on big data
- on paradigm shift on data science
- on Academia / Knowledge Graph
- on Big Tech / Google
- on Social Media
- on Legal Responsibility
- on Blockchain

Historically:

- happened when publishing started to shift to computers
- open source, the Unix industry and Linux
- Web and hyperlinking
- Libraries Cataloging - Moving online
- De facto standards vs. published standards
- US Govt as a lab
- The end of democracy

Tom Picmap was born among books. His parents were avid consumers, and by fear of missing out, made a point to buy and devore anything about a given subject.

Two kinds of data

1. Data that are collected from external sources: unknown
2. Data that are created from inside.

Techniques to handle these data.

Dodging:

- Social media
- Upper ontologies
- Google Search
- Semantic Web
    - NLP
    - Big Data
    - AI
    - Deep Learning
    - Knowledge Graphs
- Academia
- Big Corps
- Startups
- Investors
- Strategic Consulting
- Data Modeling
- XML Processing

China

China

Net state merges with Nation states. Chinese state on topi of technology

Centralized decision making. One truth, imposed on every one, whether they agree or not. No dissent allowed.

Shortens design phase, as agreement can be reached faster than with endless discussions where people express their disagreements.

Topic Maps expressing multiple perspectives. Ironic that this is what's used in China. What is it that China sees in Topic Maps that nobody sees here?

Dynamic nature of information management. Update: new kind of information, integrated naturally.

No need to go deep into the schema. Mess is more accurately defined as "adaptability" or "flexibility".

China is centralized, decision making is top-down, but it favors flexibility and dynamism.

It's both a net state and a nation state. The government is responsible for managing the technology and the corporations.

China dominates the world economy because it owns the means of production (classic Marxist theory). It is a country functioning as a corporation. This corporation has a monopoly of a lot of goods that are necessary for the supply of critical components or products in other countries: electronics, steel, clothing, etc.

Does China copy/steal US information? Many Chinese students come to the US to study and learn and do research. Research in the US is mainly conducted within the Big Tech companies. It wouldn't be surprising that there be lots of employees with Chinese origins. THer is not a lot other research in China.

For example, Topic Maps:

- I happen to have invented the concept.

- I happen to have studies how it has developed. Peak in 2008 in the West, then decrease. Rise in China, in the 2010s, with signs of activity in 2014. Seems difficult to track after due to my ignorance of the Chinese language. But I'd like to study how Topic Maps is doing, or how it is used.

Correspondence

> Throughout the book, I kept asking myself “what does this mean for Infoloom and its innovations”? The author shows how well the legal and policy dimensions of this debate are already being addressed—and Zuboff, whom he references extensively, is another example of a powerful voice who, from the social science perspectives of law and economics, is confronting these Big Tech problems.

> What you’ve done with Infoloom seems more important to at least starting to resolve the overall problem: i.e., having the ability to deliver highly customizable programs, given that AI and Deep Learning are in fact not as “customizable,” as believed, because the inner complexities make it difficult for more than a very few of programmers to get into the algorithms.

There is a chapter where he says something that is important, i.e. that a paradigm shift has happened between hypothesis-based science producing data and data-driven science.

Therefore more than customization is needed. What's needed is the return (or new shift) to a paradigm where data is the output, not the input. One of the reasons why there are so few clients aligned is that people have internalized the data driven model. There is a need to deconstruct

> If this is so, it sounds like Infoloom is providing a means for humans to regain control over the machines.

The challenge is the paradigm shift. This is why it is so difficult to sell these days. People are discouragingly not interested. In other words, they do not believe any more that it is possible to manage data, as they feel that they are managed by data. As the faith into the value of democracy fades, so goes the faith in getting any control over data. It's intimately related.

By the way, I may have been wrong on that front. I was thinking that people having to agree on the meaning of data was a threat to democracy and free/critical thinking because the agreement meant too much compromise and suppresses the nuances. Therefore I was arguing in the defense of multiple perspectives instead of a unique way of seeing things. In fact, I underestimated that the agreement is built on common, shared values, which is a precondition for anything to work. Fascism is the contrary. It's being built and promoted on the impossibility to agree on shared values. Therefore the only solution is to exert power to impose one vision over others. No compromise is possible.

Philosophically, what is happening is that relativism —the fact that anything goes— is prevailing about the search for "objective" facts, on which "traditional" science relies. The paradigm shift over data being the source of knowledge combined with "absolute relativism" creates a deadly combination, where science becomes irrelevant.

This reminds me of previous works I did in the area of history/philosophy of science and the discussions about the misuse of relativity to promote relativism. The theory of relativity is misnamed, as it is the exact opposite of relativism. I believe this discussion is central to understand what is going on now. I got these ideas from chapter 5 of the book ("What's wrong with science?")

> I haven’t seen this aspect of the problem addressed anywhere, and certainly not in Runaway Technology. Most everything in that book was about philosophy and the law. It is very “lite” on the tech (although it’s intriguing that his brother works at Google!)

I agree that it's lite on the tech. But it provides an intellectual framework that enables replacing tech in its relation with science at a conceptual level. For me, it is

the missing link and I see more clearly what I am trying to do. For that reason, it's really important.

> Does Infoloom provide the tech platform that can help deliver “accountability” to the realms of AI and ML? (For the layman, how is “accountability” defined, and how does Infoloom make this possible?)

That's a short circuit that explains the dead end. The missing part is that what Infoloom is proposing relies on the paradigm that even science and tech are accountable. As it has become painfully evident, tech is not by any means accountable. It has been made unaccountable by Big Tech. Zuboff shows it's not an accident, it's the basis of their business models. With accountability, Google and Facebook are doomed. They thrive based on the non-accountable nature of their business models. All the research and new tech developed in their labs aims at reinforcing the non-accountability.

Therefore the Infoloom proposition is not just an "improvement", it's a fight against the dominant business models. However, the fight is disproportionately unbalanced, and the investor community is also totally on the side of the big tech model. That explains why the reaction I have got recently from investors as well as big corps and the government about my proposals can be best summarized as: "Duh?" (translation: who do you think you are?)

> From my DC experience, all the talk about new, evolving laws—a la Fairfield—and new policy proposals are interesting, but it is unlikely to go anywhere. Maybe there can be progress on accountability in the EU, as Fairfield says, but we shouldn’t hold our breath for political change in the US. Therein lies endless frustration. It is like waiting for gun control: a majority of Americans know that gun control is vital, but nothing happens, decade after decade. (In fact, the best hopes for gun control are in the introduction of new technologies as well, such as biometric safeties.)

My analysis of the current situation is that the country has been through a fascist episode which is setting the stage for worst to come. Fareed Zakaria has presented

a show yesterday on CNN (https://www.cnn.com/2022/01/09/opinions/fareed-zakaria-the-fight-to-save-american-democracy-op-ed/index.html) which was quite interesting.

The movie "Don't Look Up" on Netflix, which I watched two days ago, is quite interesting too, about the impossibility for scientists to get the message through the mainstream media and the politicians and executives. Several climate experts said they recognized their own experience. In a way, I can say the same thing. It's all about the impossibility to communicate, like a glass ceiling we can't go through, or the idea that we are not living on the same planet as others.

> Ultimately, does Infoloom “disrupt” the world of AI and ML by providing highly customizable programs that are able to combine massive extraction of data with ways to efficiently locate relevant information????

It's ultimately about: who is in charge? What I am trying to tell people is that can retake control over their data if they want to. Most often the answer I get is that people don't think it's possible. It's like they have been brainwashed by the big tech ideology, supported and actively reinforced by would-be authoritarians, i.e. a combination of Republican-leaning CEOs who want to preserve their ability to make profits regardless or not it destroys the planet, or enslaves IT human operators to proprietary computer algorithms.

Yes, it's disruptive, but it's more disruptive than in any "normal" situation. When I started with IRS, they genuinely wanted to solve a problem to improve the trust taxpayers could place in them, to make the system work better. I don't think anybody currently in government still has this attitude. They have already lost control, they are reactive and not proactive. They don't believe any more that human thinking can be powerful enough to modify the way power is exerted. It requires much more than just a technological solution to make them open to this kind of solution. Even it the planned outcome is the end of the American experiment, as we have known it.

> Sounds important. Because Infoloom can contribute so much, it remains vital to get the core mission clearly and strongly stated. And such a mission statement

> is pretty different from: “We provide data curation solutions with custom integration and interactive tools to help you make your data what it should be, instead of what it happens to be.”

Agreed. But it's possible that what Infoloom is bringing is not what most people want to hear. The challenge is to find people who are already aware of that kind of issue. My hope is that even if it's marginal, there are a number of initiatives that go in that direction. For example, local initiatives such as in New York City: https://www1.nyc.gov/assets/cto/#/ or the Center for Humane Technology and satellites. FYI, Roger McNamee is one of the founders of the Center for Humane Technology. So, there may be a handful of investors who are aware of what's at stake. The question therefore may be eventually: How to present them with something that makes sense?

> Moreover, it seems that Infoloom might be able to underscore the importance of this problem—and of the solution—by showing what’s up in China. That’s worth exploring, and why we should have phone conversations/Q&A with these AI/PRC experts.

Yes, China is a good (counter-)example of what can (and should not) be done. The interesting part about it is the fusion between the interests of a nation state with the big tech business models. Facebook naturally tends to nurture Chinese-like capitalist/communist dictatorships. If Big Tech companies continue to refuse any kind of accountability (which is honored both by the Republicans but also by the Democrats), it would be no surprise that US becomes a sort of sub-China kind of regime, with a slight difference. In China, the government controls the technology, and they have the resources and abilities to do so. In the US, government has renounced to control Big Tech, where profit is the engine that powers the economy. The government makes sure that they let Big Tech do as they please, even if it means that the "people" are completely left aside. It means inequities going to the roof, increased poverty, and return to a kind of slavery (where the people are denied the right to vote). The Chinese model is more efficient.

The proposal for a NIT based on the NIH aims at showing that it there is a precedent of government-funded research in the public interest, coexisting with private research, that ensures that US is in a leading role, worldwide, on medical research. It could be playing a similar role on technology research, provided that the taboo of letting Big Tech do as they please is broken. In a way, that illustrates the core challenge that we are facing right now.

2022-03-31

Skeleton

Chapter 1. It doesn't have to be this way.

- Heading in the wrong direction.
- Tech in this context. Net states, Data Extracted from customers, Impact on law.
- Technology Ideology: inevitable, any opponent is a Luddite
- Purpose: unravel the ideology, consider it a fight, resistance needed against mainstream.
- Focus: data, critical analysis of big data, AI and ML
- Big data is big because lots of data is extracted. Not all is necessary.
- Smaller amounts still seem overwhelming:
    - photos in a smartphone,
    - items on a government web site
    - items in an online seller catalog
- Big data processed by algorithms as black boxes
- No transparency + magic power => blind trust, with sense of powerlessness and suppression of freedom of expression.

Chapter 3

- Knowledge is layers upon layers. Referencing, quotations. External layers.
- Metadata vs data: One's metadata is someone else's data.
- Information can be messy. Several names, languages, schemas.

- Multiple perspectives possible: two indexers would produce slightly different indexes, both good.
- Human brains operates by associating thoughts. Somewhat imprevisible.
- Human communication pitfalls. One person's understanding may differ from what another person is saying.
- Nothing obliges that sharing with groups of families and friends, exchanging email messages, maintaining a calendar, getting involved in a group of friend, reading a book is automatically data acquired by the company providing the service.
- Nothing obliges that algorithms have to be kept secret. Page ranking could be documented. Algorithms could be open sourced, or at least shared by their users.