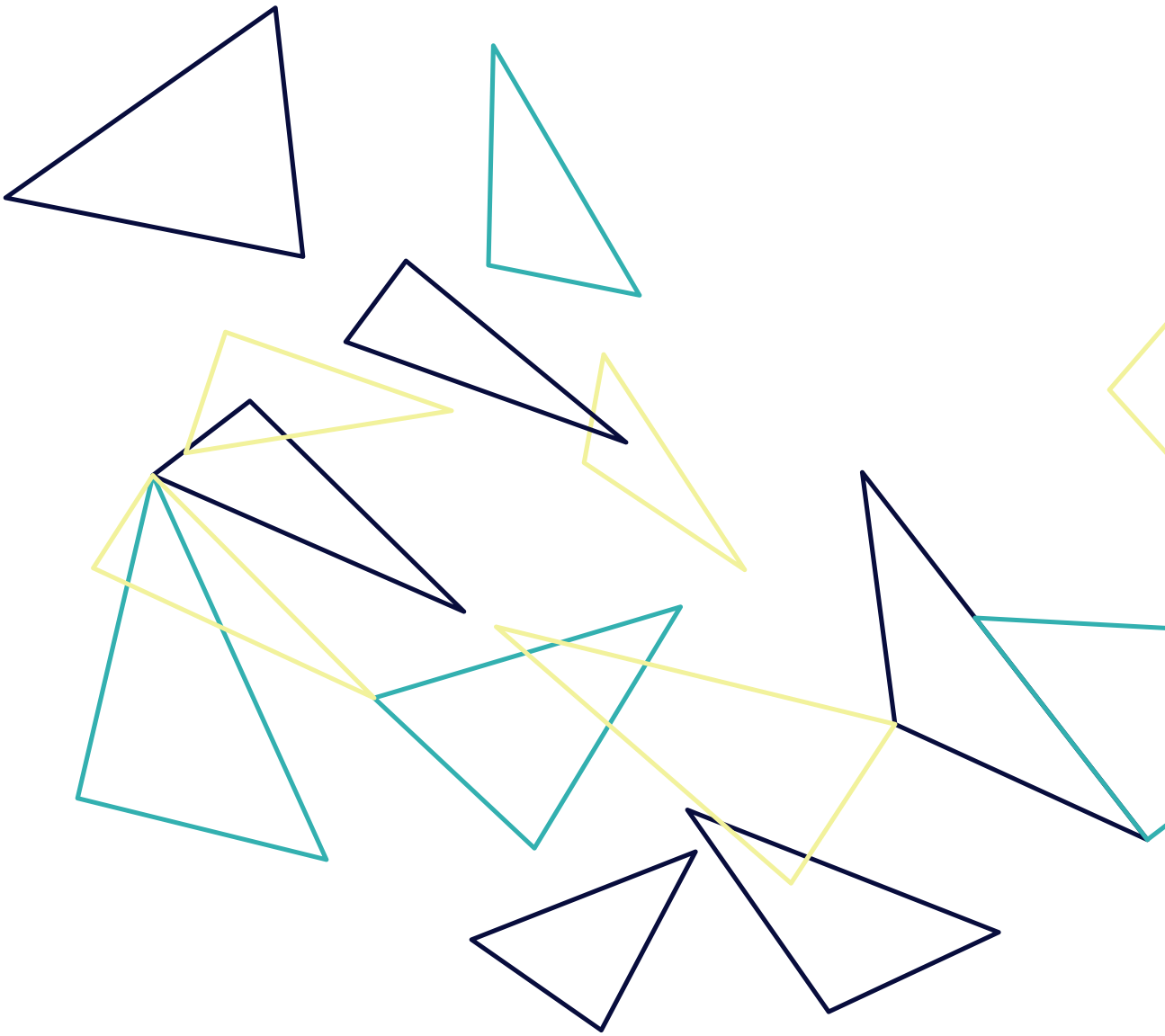
Ethics & Tech

2019

OPTIC

(RE)BUILDING TRUST IN TECHNOLOGY



with the support of :

international foundation

Foreword

Fr. Eric SALOBIR, op, President of OPTIC

Something has changed.

Over the past year, scandals and media revelations have gradually chipped away at the public's trust in technology. Despite all its dazzling qualities, it seems to have revealed a much darker side. Has tech lost its charm? In the last month, more than half of tweets mentioning artificial intelligence were about "deepfakes", technology that makes it possible to create extremely realistic-looking fake videos of someone saying something they never actually said. Furthermore, fifty-five percent of these conversations were marked by fear. Ironically, though, these numbers did not come from an opinion poll, but instead were provided by… artificial intelligence software that scans the internet to pick up trends and weak signals in our language. In short, technology is diagnosing the problems that technology creates.

Does this make tech part of the solution as well as part of the problem? Are people not destined to be made obsolete by the all-powerful machine if we can just manage to resist the temptation of replacing them and instead opt for fruitful collaboration? Once again, we need to rethink our environment and come up with new, technology-enabled methods. Blockchain is disrupting our relationship with information. Artificial intelligence is changing the nature of armed conflict; deploying digital technology in sovereign activities is transforming citizens' relationship to the state to the point of challenging the status of nations; brain-machine interfaces promise treatments for neurodegenerative diseases, but also cause us to question what makes our personality or even our identity.

How can we restore trust in technology in light of these challenges? Or rather, how can we ensure that it is actually worthy of our trust, that innovation is always a component of progress?

These questions are the motivation behind this report, which should not be seen as a monolith, but rather more as an impressionist painting to which personalities as diverse as UNESCO Director-General Audrey Azoulay and astronaut Bertrand Piccard have offered to add their color. In their own way, each of them will shed light on a question and encourage us to join the discussion. It is my pleasure to invite you into the conversation and to share your reactions with the members of the worldwide OPTIC network. If we want technology to become a tool for a better society, everyone is going to have to lend their voice.

Happy reading!

"Overcoming fear by directing new technologies toward the common good"

Audrey AZOULAY, Director-General of UNESCO

From the controlled use of fire to the harnessing of nuclear energy, and from the invention of the printing press to the invention of the steam engine, each one of humanity's technical achievements has given rise to its own set of hopes and concerns. Digital technology and artificial intelligence (AI) are no exception, and we must work to ensure that the new possibilities afforded by technological progress serve the common good.

Our collective imagination is marked by a wealth of cultural production – movies, books, all kinds of dystopias that reflect our fear of being overtaken by our own creations. Stephen Hawking himself predicted that artificial intelligence could destroy the world.

Today, we find ourselves at a point where we need to articulate, on the one hand, the radical nature of choices regarding technological, social, and economic innovation, and, on the other hand, the responsibility associated with ethical choices.

This is not about being afraid or being naïve; it is about being fully aware of our responsibility.

This responsibility is twofold. We must understand what is at stake, and we must define the path which will allow us to use artificial intelligence so that it serves the common good, as collectively established in the United Nations' 2030 Agenda for Sustainable Development.

The fears we harbour find fertile ground in the unanimous belief that with artificial intelligence – along with other emerging technological advances such as robotics, big data analytics, and the Internet of things – humanity enters a new era and faces the unknown. Artificial intelligence raises questions not about technology, but about our own humanity. It raises questions about politics, philosophy and ethics.

What we are experiencing is a veritable anthropological revolution. It affects every aspect of our lives – our relationship to work, to time, to space, to others, to the human.

That is why this technological revolution compels us to ask questions. By choosing the way in which we develop these technologies – how we control them, how we direct them – we are choosing how we will forge the world of tomorrow.

Already, the innovations achieved thus far expose us to situations that make us wonder. For example, the use of self-driving cars raises questions about how we determine responsibility in the event of an accident, and about which criteria should be applied when programming a decision-making process involving a potentially fatal choice.

Certain medical diagnosis systems using artificial intelligence have proven to be fast and reliable, but again, in the event of an error, who is responsible? What biases are built into the algorithms?

We must carefully consider the matter of responsibility, which is ultimately ours since we are handing

over this kind of power to machines.

Before trying to understand the limits of artificial intelligence and, indeed, whether they exist, before even envisaging robust and self-aware artificial intelligence, we must first answer the fundamental ethical questions raised by current AI technology: Who decides on the priorities and values that are programmed into a machine's algorithms? What limits should we set for a machine's independence and decision-making power?

In a world where access to these new technologies is highly unequal, how can we make sure that they do not widen the development gap between countries and between genders? Given that deep learning is based on historical data, how can we make sure that the decisions it produces do not aggravate past biases? How can we make sure that the power and information offered by artificial intelligence are not used as tools of oppression?

Inequality, the abuse of power, discrimination – all the ills that we fear might be fuelled and strengthened by new technology – existed before these technologies came along. These misuses, these abuses, these inequalities are our own. The question is thus the following: how can we make sure that new technologies do not reproduce or exacerbate our own failings and ensure, instead, that these technologies serve to strengthen the common good in the interests of humanistic values and human dignity?

Wonderful potential tools for the Sustainable Development Goals of the 2030 Agenda

New technologies also open up unprecedented opportunities for the development of societies, of knowledge, and of human progress.

They can be tools for resolving some of society's most crucial issues.

I am thinking, for example, of certain inspiring projects from around the world, which have been presented at UNESCO. These projects seek to employ new technologies in a vast range of contexts, including biodiversity monitoring in tropical forests, the development of sustainable agriculture in Africa, the fight to end genital mutilation and domestic violence against women, helping deaf and hard of hearing persons to experience music, and the personalization of education through the real-time analysis of student learning.

We need to reflect and act collectively to ensure that new technologies shall always serve sustainable development and the common good.

UNESCO's role as a laboratory of ideas.

UNESCO is particularly well placed to support the debate on new technologies. This advantage stems, first and foremost, from its universal mission within the multilateral system of the United Nations.

UNESCO is a gateway between its Member States and civil society, the technical and scientific community, the university sphere and the private sector. It offers all these actors a platform for discussion and debate.

The opportunities offered by new technologies, and by artificial intelligence in particular, are radically transforming all the fields within the purview of the Organization's mandate: science, education, culture, communication and information. Thanks to its multidisciplinary expertise, UNESCO has full understanding of the issues at stake.

UNESCO has thus been fulfilling its role as a laboratory of ideas with regard to artificial intelligence. A series of AI-themed meetings was held in Paris in

2018. One, for example, involved a discussion entitled "Artificial Intelligence for Human Rights and SDGs: Fostering Multi-Stakeholder, Inclusive and Open Approaches" and took place within the framework of the Internet Governance Forum (November 2018). Furthermore, the first regional conference on artificial intelligence for development in Africa was held in Morocco in December 2018. The subject is a universal one, and UNESCO is equipped to support reflection upon it throughout the world and on a variety of scales.

In March 2019, UNESCO furthered the conversation with the first international conference at UNESCO entitled" Principles for AI: Towards a Humanistic Approach?"

Reducing the inequalities in access to AI

Our first challenge is to reduce the inequalities associated with access to artificial intelligence and to new technologies.

At the 38th session of the General Conference, the Member States of UNESCO adopted the four principles for Internet universality. The ROAM principles are: human rights, openness, accessibility for all, and multi-stakeholder participation. Our actions regarding the Internet and artificial intelligence must be founded on these principles.

These principles are at the core of our programmes to teach coding to girls and young women in Africa, at the intersection of Global Priority Africa and Global Priority Gender Equality, and of the promotion of women in science. They underlie AI training courses organized in partnership with the private sector. Internet universality is also a central factor in the promotion of open science, open innovation, and universal access to knowledge, which make it possible to close the gaps in technological progress between countries.

Questions raised throughout all sectors

While it is essential to ensure that the requisite skills are held by as many people as possible in order to harness the multiple possibilities offered by AI, we also need to develop a critical view on its consequences.

Artificial intelligence has consequences for all domains, which need to be considered and addressed. It is changing experimentation and explanation methods in the social and natural sciences and it has an impact on the reasoning applied. AI-based artistic creation calls into question the status of the author, while recommendation algorithms threaten the preservation and promotion of cultural diversity.

In the field of communication and information, artificial intelligence raises hopes for the bolstering of quality journalism and the filtering of hate speech. At the same time, however, it raises concerns about freedom of expression and could increase the spread of disinformation if its development is not based on human rights and informed by multiparty engagement. UNESCO develops media literacy programmes to raise awareness of such risks and of algorithms.

The use of artificial intelligence in the various sectors

In all our areas of action and for all our objectives, new technologies also provide opportunities and positive applications which we seek to identify and put to good use with our network of public and private partners.

An international conference has recently taken place in Beijing. Its participants will be studying the possibilities of artificial intelligence in the realm of education, the accessibility of learning, the personalization of learning pathways, and so forth. Artificial intelligence

has become an essential factor that must be taken into account in order to achieve the goal of quality education for all.

We are thus working with Microsoft to ensure that the decisions taken with regard to the role of artificial intelligence in education are well informed.

Artificial intelligence has countless applications. By studying images taken by drones, for example, we can fight chimpanzee poaching in the Mount Nimba Strict Nature Reserve in Guinea.

We also use drone imagery and its processing by AI in the domain of culture: we have partnered with Iconem, a company which produces such images, including, notably, images of endangered sites such as Mosul. Artificial intelligence and new technologies play an important role in UNESCO's Revive the Spirit of Mosul initiative, whereby the Organization is coordinating international efforts to restore and rehabilitate cultural heritage and revive educational and cultural institutions in Iraq.

The need to orient new technologies toward the common good.

The mere evocation of these prospects is inspiring and generates confidence about future opportunities. Artificial intelligence, however, is only as effective as the project in which it is used, and it is contingent on the conditions defined by those who have programmed it, based on the data they have given it.

We must make the relevant choices with eyes wide open, if we do not wish to witness "the end of the Enlightenment", as expressed with concern by the youthful 95-year-old Henry Kissinger.

New technologies have the value that our use of them confers. This value can be immense.

Rebuilding trust in new technologies supposes an awareness of the choices they imply. It requires that we take measures to ensure that when we develop these technologies, we do so for the common good, ensuring that everyone everywhere benefits from this technological revolution, which knows no borders.

For this, it is essential that we reflect on the implications of these new technologies, and that we establish public policies to guide their use and ensure ethical principles are applied to them.

UNESCO's standard-setting role: expertise and decision-making with regard to AI.

For over 25 years, UNESCO has been developing unparalleled expertise on the subject and has been working on ethical issues in science and technology, including matters as decisive as the human genome, genetic data and climate change. Its standard-setting role allows the Organization to adopt regulations that are binding on its Member States.

Through the World Commission on the Ethics of Scientific Knowledge and Technology (COMEST), UNESCO has already carried out important work concerning the Internet of things.

UNESCO is thus preparing to fulfil its role in the sphere of artificial intelligence and its ethical questions.

A preliminary report prepared by COMEST was presented last month to the 57 Member States of the Executive Board of UNESCO. The report underscored the validity of the Organization's mandate, in support of which the objective was set to develop guiding principles for the ethics of artificial intelligence, in the form of a UNESCO recommendation.

This item is now on the agenda of the fortieth session of the General Conference of UNESCO. This is the first stage in a process which will extend to 2021 and the aim of which is the development of a new

standard-setting instrument, following a worldwide conversation enabling greater awareness of the matter.

Conclusion: collectively building trust

Through its action, this reflective process and the development of a standard-setting instrument, UNESCO seeks to guide the use of new technologies so that they contribute to peace and sustainable development, for the benefit of all.

To achieve this, we must define the principles and main directions of the development of new technologies. We must collectively build engagement on the part of all stakeholders, governments, private-sector players, the scientific community and civil society in order to ensure that this undertaking respects basic human rights and contributes to sustainable development.

We owe this to present and future generations.

Introduction

Pierre GUEYDIER, Director of Research, OPTIC

It is with great pleasure that we present to you the first edition of the OPTIC network's ETHICS & TECH report. It is the product of many months of rich collaboration by contributors and authors from diverse geographic and intellectual backgrounds with the aim of providing an accessible overview of the collective and contemporary issues of digital technology – be they human, political, or anthropological – to anyone seeking to better understand them.

These issues are rife with uncertainty and complexity, which in turn erode trust. Recent months, though, have borne witness to a fundamental shift: The digital utopia is gone for good, not just for a handful of individual thinkers but for large swaths of public opinion and policymakers. The risk, and the simplest option, is that of designating scapegoats, which are all the easier to find since they have been built up to the pinnacle of the progressivist digital utopia over two decades.

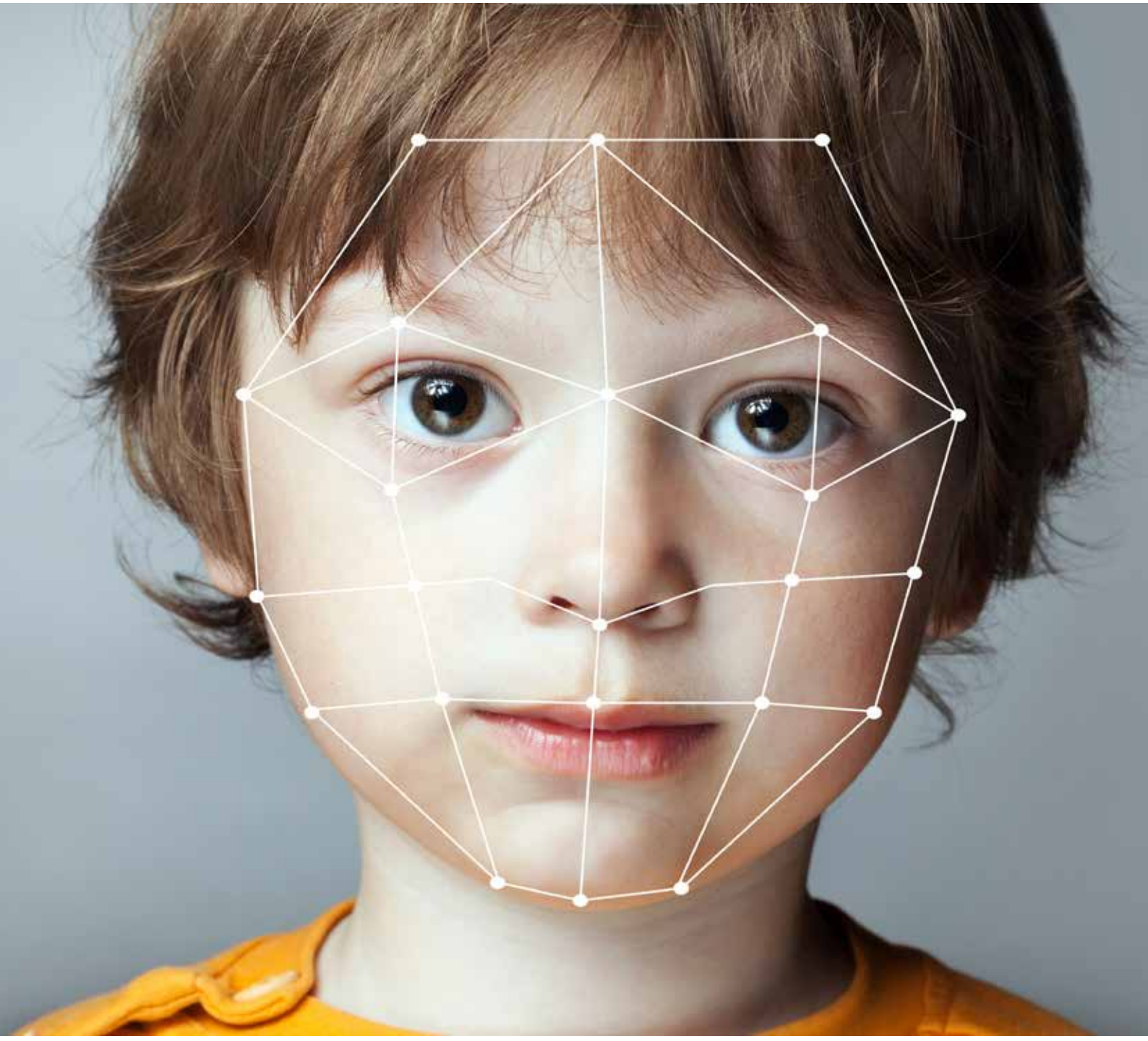
The general focus of this report is not to submit yet another indictment against those actors: We are all collectively responsible for the negative externalities of how we use digital technology.
Our aim is, without downplaying the issues, to go beyond the atmosphere of catastrophism and propose some hopeful possibilities to restore trust in technology.

We are facing a crisis of progress, that political and social driver of humanity for 250 years running. A pivotal moment has come where our awareness must collectively seek out new types of progress based on the common good. Otherwise, progressivist ideology will find itself in a dead-end loop of producing ever more inequality, which inevitably leads to social violence.

To reorient the situation and discussions on these lines, we have done our best to provide a clear, accurate explanation of these phenomena to proffer food for thought on matters such as human dignity through the concept of data, political participation/representation, the possible revival of mutualism with blockchain, digital peace, and the ethics of responsibility through the virtue of frugality.
Our report is divided into four parts.

To begin, we will explore the relationship between people, data, and machines. We will then cover shifts in governance and the social contract. In a third part, we will focus on two essential and paradoxical disruptions: blockchain and trust, then the role of AI in war. Finally, we will open the floor and invite readers to reflect on the ethics of innovation in terms of sustainability, technological frugality, and innovation "by design".

Contents



ETHICS & TECH Report 2019

Part I
PEOPLE, DATA, MACHINES

In the first part of this report, we aim to shed light on the most recent debates on "data", a concept that manages to be centrally important but also underdiscussed and easily misleading. Its central importance stems from the fact that the basis of any artificial intelligence system relies on vast quantities of information to feed its learning algorithms. It is underdiscussed because harvested data is often seen as a stand-in for reality, appearing neutral and rational. After all, what could be closer to reality than information collected automatically by mechanisms free from error and ideology?

One of the main misconceptions to come from the current technology backlash can, in our opinion, be traced back to the wooly definition inherent to the notion of "data."
To attempt to explain this, we will be placing the production of data in the digital age in the context of a broader movement in science and technology that is based on the act of writing down data, which itself has joined forces with another, equally important phenomenon in the modern era: the bureaucratization of human political and economic organizations. This report also discusses the most advanced stage of data intrusion using the concept of "nudge", which combines mass production of data, choice architecture, and social engineering. We will then examine what the ubiquitous invocation of data means as well as the almost magical qualities lawmakers and policymakers give it. To clarify and anticipate the consequences of data and machines on people, we will explore the burgeoning field of neuroscience and the latest major advances in brain-machine interfaces. For informational purposes, we will provide an update on the state of the art of imaging technology and invasive cerebral interfaces. As a complement, we will present the main ethical problems of these technologies as already stipulated by law, albeit only as relates to experiments.

In conclusion, to round out the didactic and critical angles related to the people-data-machine triad, the final contribution will explore a hypothesis for restoring trust after the brutal appropriation of our personal data by a few monopolistic actors within the digital economy. By centering the debate and any practical solutions on the notion of the commons, making data portable across social networks could avoid the appropriation of the social ties our data reveals.

Contributors:

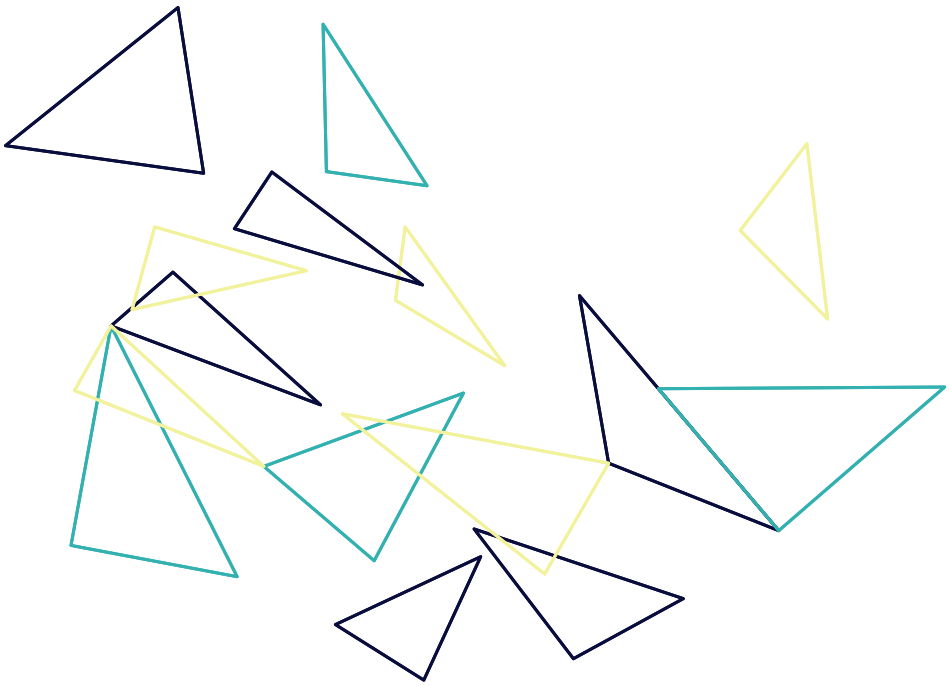
Pierre **GUEYDIER**

Lionel **MAUREL**

Claire **SOMERVILLE**

Laure **TABOUY**

Bernardas **VERBICKAS**



CONTENTS

Data work, nudge, and reductionism

Pierre GUEYDIER

Big Data, open data, raw data, data-driven… The term "data" is everywhere in digital technology debates. This paper seeks to question the misleading nature of this seemingly neutral term. One of the main problems with assessing the effects of digital technologies probably lies in the extreme difficulty of defining the very concept of "data", which connotes a certain naturalness that belies the existence of the processes and players, often hidden and erased, that created it.

Science, technology, and inscription[1]

Day to day, there is a very strong relationship between scientific and technological production and the act of writing in its many forms. Interest in the act of writing within the edifices of science and technology saw new life in the 1980s. We thought an ethnography, so to speak, of the work associated with scientific and technological production would aid us in shedding light on the mystery surrounding the concept of "data".

Following in the footsteps of Derrida, these works[2] shifted attention to focus solely on semantic content and divulge the active and material aspects of its production. As a result, scientific and technological literature cannot be thought of as a mere vessel, neutral and transparent, but rather as a participant via its composition, organization, and the way it creates knowledge.

At the very heart of the issue, which also includes the idea of data in the digital world, is the matter of knowing how scientists write down their findings to reflect reality. Simply put, it is impossible to isolate knowledge of any kind, even when reducing it to a unidimensional "datum", without paying equal attention to the material factors that allowed it to be inscribed. This especially applies to methods of displaying data, by which phenomena originating in a "natural" setting are made visible. The reality of these phenomena is progressively translated into inscriptions that can be seen and interpreted, either as written text or graphical representations (map, diagram, table, etc.). In science, data has "written" properties that make it coherent and allows it to be spread[3].

By expanding the idea of text to include the broader idea of graphical "inscriptions" (e.g. drawings, graphs,

1. This chapter on data is largely based on the work of Jérôme Denis at the Centre de Sociologie at the École des Mines in Paris, especially his paper *Le travail invisible des données*, Presses des Mines, 2018.

2. Shapin, S., Schaffer, S., 1985. Leviathan and the Air-Pump, Hobbes, Boyle, and the Experimental Life, Princeton University Press. Latour, B., Woolgar. S., 1988, La vie de laboratoire, Paris, La Découverte.

3. Denis. J. Op. cit., p. 33.

recorded numbers, points), one must look not just at the human authors of scientific literature – researchers or technicians – but also at the "inscription devices", which contribute a denser description of reality. In a laboratory, from the sensor recording a variable to the publication of a paper in a scientific journal or the result of an algorithm, there is always an extensive chain of reading and inscribing. As is the case today with digital data, one must be able to trace back its long, written history to understand the role of writing, its process, and the skills required to create it.

To follow what science and technology studies have been teaching since the 1980s, it is vital to be aware of the fact that the digital world is above all, perhaps even more so than laboratories, a writing environment filled with recording instruments (those ubiquitous sensors and devices) coordinated by an immense field of writing (scripts, codes, protocols). Today, this dynamic is invisible and creates a massive "black box" effect. This opacity contaminates every level of society and is beginning to produce widespread damage to public trust in these intrusive methods.

Data workers

Ethnologists observing activity in the laboratory have pointed out another aspect vital to our purposes: the invisible, hidden effort and skill needed when recording on a large scale as well as when organizing and translating reality. To go from one inscription to another, it is crucial to examine all of these "negligible" tasks done by whole hordes of data workers as a means of shining a light[4] on all of the "little hands" doing the "grunt work".

The positivist and reductionist roots of the digital age can generally be traced back to the 1840s and the ubiquitous, normalized, and mechanized written communication of bureaucracies still in their infancy. Between the 19th and 20th centuries, public institutions and businesses had a wide array of writing technology and infrastructure. Then, just as the civil service developed a public system of population statistics, businesses experienced a management revolution by rethinking the market, which was endowed with writing technologies for measuring, calculating, and sequencing. Even before the digital age, data became a commodity for the public and private sectors as a key component for coordinating all types of exchange. At this stage, it is necessary to point out the political dimension of writing processes by examining the value placed on the various stages of data production. The mechanized normalization of producing writing and data within bureaucratic organizations is primarily centered on the political principle of efficiency. By valuing efficiency, data work and data workers are, often imperceptibly, relegated to the shadows cast by algorithms and machines. This is spectacularly true when it comes to the discourse around and perception of artificial intelligence, in which almost no credit is given to the colossal efforts made by AI "handlers" and the microwork done by people in the shadows[5]. In fact, at the heart of even the most seemingly autonomous processes, there is still a proportion of essential work done by people on the edges and in the interstices of the network, including maintaining the machines used to produce and spread data.

There are in fact many cases that show that, although

4. Star, S. L., 1999, "The ethnography of Infrastructure", American Behavioral Scientist, vol. 43 (3), p. 377-391, cited by Denis, J., op.cit. p. 45.

5. Casilli, A., 2019, *En attendant les robots, enquête sur le travail du clic*, Edition du Seuil.

producing masses of data may look simple, automatic, "mindless", and valueless, a closer look reveals a density and complexity worthy of our full attention[6]. Data workers, a term that includes more or less anyone who has contributed even queries to a search engine, are part of a massive "back-officing" of the world. Microtasks, coordinated by ever more monopolistic platforms are the harbingers of a post-capitalism with such negative and violent social externalities that even a sovereign state will struggle to counteract – if it is not doing so already.

A new type of data: "raw data"

Of the many varieties of data, it is "raw data" that should receive our full attention here because the historical status of data has changed along with its ubiquity in any discussion of digital technology. The massive liberation and increased speed of data distribution has become an easy stand-in for transparency, innovation, democracy, and efficiency to the point of becoming the forefront of the vast solutionism movement[7] so typical of discourse and agendas in the digital age.

In this worldwide technophile movement, there is one term that catches the eye because it is found within the positivist aspirations of champions of so-called "open" systems, free software, the bureaucratic virtue of transparency, access to information, and Anglo-Saxon accountability. The term in question is "open data" and its corollary of "raw" or "unmodified" data.

This previously unknown data entity, appearing first in 2007 at the meeting in Sebastopol that laid the foundations of open data and having since been amplified by the biggest names in the digital transformation[8], would go on to create "raw data" as a new type of information, one that does not refer to files created by bureaucratic administration, nor to statistics. Instead, it refers to a type of information that is a more fundamental precursor to the usual categories – with no further definition. It refers to something that is "already there", something that pre-dates any type of write-up and that would be easy and straightforward to "liberate". It is a theory of information that runs counter to what this paper has already pointed out: the real and material significance of data production, processing, and distribution.

The idea of "raw data" aims to dematerialize the concept of data, or even to naturalize it by granting it the status of a raw material and commodity. This neo-positivist ideology, however, does not bear out in reality and, in our opinion, contributes to an oversimplification of digital data, especially when it does not measure the political aspect of its social fabrication in the positive and collective sense of the word. In doing so, this conception of the idea of data, which we deem false, has major consequences when it implies to the public that gaps in the use of so-called "raw" administrative data are suspected of feeding a binary opposition between transparency and opacity. Indeed, the perfect datum, innate and discovered naturally by "platformized", crowdsourced programs, does not exist and runs counter to the reality of the discrete, complex mechanisms that create it. We must therefore abandon any realist position and admit that data is not an informatic entity that already exists and just needs to be disseminated (or "liberated"), but the pro-

6. Denis, Jérôme, op. cit., p. 97.

7. Morozov, E., 2014, *Pour tout résoudre, cliquez ici, L'aberration du solutionnisme technologique,* FYP.

8. From Denis, J. op.cit., p. 153: In the words of economist Rufus Pollock, "give us the data raw, give us the data now", or Tim Berners-Lee, co-inventor of the internet, "we want data raw".

visional result of a delicate process of creation. As Latour once wrote, we must admit that data is always something that has been obtained[9].

Nudge and the architects of choice

Encouraging people to change their individual or collective behavior while maximizing cost effectiveness (be that cost financial or political) is the ultimate goal of any human government. Whether it is the public authority acting as part of a biopolitical vision to protect and develop the population (health care policy, security, ecology, etc.) or a business, whose reason for existence is managerial, productive, or commercial efficacy, human organizations all seek progressivist "change".

Until recently, this mission consisted of devising a variety of top-down incentives that were always limited by the risk of excessive repression or authoritarianism (costly and counterproductive), long lag times (between decision making, implementation, and measuring the effects), or even time- and space-limited effectiveness.

Over the past decade, however, the aforementioned increase in data work, social psychology, management, and the digital platformization of social relationships have given rise to a general theory of gentler influence now known as "nudge". While it is not yet a household term, we believe it will become a central topic in the coming months and years. As was the case for marketing and advertising, a democratization of these behavioral techniques is essential.

Encouragement or gentle discipline?

The founding work on this trend, written by Richard Thaler[10] and Cass Sunstein, was published in 2008[11]. In the introduction to their book, the authors explain that economism and its notion of the rational consumer is pure fiction. Instead, the many real examples of social dysfunction – obesity, debt, and lack of social security are examples given by the authors that this paper will discuss – give credence to the notion that the idealist perception of rational human behavior has failed. *Homo economicus*, whom Thaler and Sunstein cleverly nickname the "Econ", makes judgement errors every day revealing false reasoning and multiplying biases. Two areas of behavior define the power of nudge: inertia and the possibility to use it to design choice architecture. For example, in a self-service cafeteria, it is very easy to direct people's choices simply by displaying the food in a certain way.

This simple method of incentives through choice architecture, like the artful displaying of wares that has been done since the dawn of trade, can be drastically upscaled with digital-age trading platformization. Very quickly – by the late aughts[12] – nudge theory was being studied for its uses in policy. In highly liberalized American and British politics, nudge theory provided a theoretical third way between state interventionism and ultraliberalism. When it came to health care and high debt, supporters of nudging claimed to have found a solution for counteracting behavior considered to be antisocial while maintaining a lack of state intervention. Anglo-Saxon public policy has always erred on the side of *laissez faire*, a liberal policy whose

9. Denis, J., op.cit., p. 178.

10. Who won the 2017 Nobel Prize in Economics for his work.

11. Thaler, R. H., Sunstein C. R., 2008, Nudge : Improving Decisions About Health, Wealth and Happiness, Yale University Press.

12. In 2008, UK Prime Minister David Cameron added a "Nudge Unit" to his cabinet, following Barack Obama's example.

hypothesis postulates that everyone can be their own entrepreneur and consistently make the right decision motivated by self-defense and their own calculated self-interest. *Laissez faire* also rejects any type of coercion, in keeping with liberal doctrine, whose goal is to limit state meddling in individual lives as much as possible.

Thaler and Sunstein go on to counter these two postulates but do so using an original approach: For them, it is necessary to accept that people are fallible and to consider instead that, given the complex set of choices available when it comes to buying health insurance in America, for example, help is indispensable to navigate what is on offer. Moreover, this type of state paternalism is not the same as coercion; skillfully employing choice architecture, interface design, and optimized data processing could make the external nudges on choice imperceptible to users. To this end, Thaler and Sunstein coin an almost Orwellian oxymoron: "libertarian paternalism".

Choice architecture, which seeks to do no more and no less than improve the lives of users of public services without their knowing, is naturally underpinned by an eminently political vision of social relations. It is therefore symptomatic that the primary areas of application for nudge (obesity, debt, lack of insurance) are typical stigmas of poverty. By postulating a hypothesis based solely on the behavioral origins of these negative social traits, however, the authors completely erase the essential social and political dimensions of these inequalities. Their book and their approach make no mention of the collective political responsibility for these issues or their remedy.

Although the success of nudge in the late 2000s was linked to both David Cameron's conservatism and the health-care debate under Barack Obama, pairing it with data turn and artificial intelligence may well have an enormous impact on globalized choice architecture policy, especially digital platforms and players that are deliberately trying to achieve or maintain a monopoly in the area of information processing.

Big brother is nudging you!

Of course when it comes to guiding students to make healthy choices in the cafeteria, reducing speeding with automatic radars that use smiley faces instead of words, or helping people to not forget to renew their insurance policy, everyone can agree on a general application of this type of choice architecture and "libertarian paternalism".

The road to hell, though, is paved with good intentions and below are two quite well documented examples of ethically questionable usage of these social neuro-engineering techniques.

One of the most promising applications of nudge is in the relationship between humans and robots in the form of an extension of choice architecture and interfaces. Building emotional relationships with robots has become a field of research unto itself which starts by detecting, then classifying, and ultimately modelling emotions using verbal and non-verbal cues. From commercially available voice assistants to the potential for robot carers of the sick or disabled, nudge theory has a wide field of application to study and improve certain groups' empathy towards machines. Modeling and implementing language-related social skills such as politeness, humor, or irony is what some research teams are focusing on to identify and interpret certain behavioral cues by human users that indicate social and emotional interaction in order to then use humor to engage the human user in a long-term

relationship with a Nao robot[13]. Chatbots' and anthropomorphic robots' ability to detect, interpret, and simulate emotions – made possible by deep-learning and AI – means that they can emotionally profile people in real time.

Once an emotional state has been determined, the machine can calculate the target (increasing emotional well-being in the elderly, for example) and use conversational nudge, such as humor, to establish and have a dialogue with the person to guide them towards this type of behavioral objective. Emotional attachment to a machine is the behavioral objective in this case.

However, these influencing methods drift very quickly into conditioning, especially with young children. An example of this is when a voice assistant teaches a child to be polite to it using libertarian paternalism, which translates here to a spoken message of reward when the child addresses it politely ("please", "thank you").

The acceleration and scale at which nudge is being implemented through digital means paired with artificial intelligence has not failed to raise recent interest in one particular area of the public sector: elections. A textbook example of choice and democratic free will, voting presents itself as a natural area of application for nudge. Ever since Barack Obama's 2008 campaign, voter data has been a strategic pillar. Profiling and mapping were prime drivers of his massive and famously successful canvassing campaign.

Eight years later, with the development of nudge theory and the aid of advances in processing Big Data, the 2016 American presidential campaign employed "big nudge" or "hypernudge"[14]. Voters were openly profiled by their fears (immigration, gun control), which, according to campaign advisor Roger Stone, are the most powerful drivers[15]. Targeting these demographics by their location in key counties in swing states coupled with massive social-media campaigns containing bold-faced lies to trigger fear in those voters turned out to be a remarkably effective choice architecture using nudge that was unprecedented in its efficacy and precision.

Reductionism

Invisible data work and the development of nudge theory underline the importance of an age-old school of thought that is becoming ever more relevant. This philosophical concept has been a driving force behind technical progressivism since the 18th century and seems to us to be worth revisiting to have an overall view of the issues raised when digital technology intrudes in all aspects of individual and collective lives.

This powerful philosophy is called "reductionism". Notwithstanding its sizeable role in scientific and technical efficacy, our opinion is that it cannot be allowed to extend to the dogma of "reducing" human complexity down to data, even masses of it, that can be modeled and manipulated.

Reductionism is a pivotal concept in Cartesian mate-

13. https://lejournal.cnrs.fr/billets/rire-avec-les-robots-pour-mieux-vivre-avec

14. Yeun, K, 2017, "'Hypernudge': Big Data as a mode of regulation by design", *Information, Communication & Society*, Volume 20, 2017 - Issue 1: The Social Power of Algorithms. https://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1186713

15. Watch Dylan Bank and Daniel DiMauro's excellent documentary "Get me Roger Stone", 2017. https://www.youtube.com/watch?v=5lPyv4KgTAA

rialism, which aims to simplify existing phenomena down to as many elementary components as necessary. In materialist doctrine, all that exists is matter and physics is the fundamental science. The resultant analytical method, essential for scientific processes, has demonstrated its full worth and has been endorsed by the biggest names in science[16]. However, the absolutionist take on this principle, by which only that which is composed of matter, physical phenomena or anything that falls under measurable "data" should be considered to exist, is of course questionable.

Indeed, as the title of Pablo Jensen's book[17] states, society does not fit into equations. Physicism, by which physics is the general model that explains the material world, cannot be epistemologically transposed onto issues that are sociological, anthropological, or especially political in nature. Rationalism's key ideas, such as reproducibility and predictability, clash with the complexity of human relations. Statistics, on the other hand, is always defined as a rational science of social affairs that could experience an epiphany thanks to the digital turn. Political excesses in social engineering result in the blatantly misguided belief that any social issue (work, health, violence) can be modeled by isolating and simplifying it. The repeated – even consistent – failures in economic forecasting underline how excluding certain effects to simplify models or confusing statistical correlation[18] with statistical proof cause the reductionist approach to social phenomena to fail.

There are two opposing prediction models: extrapolating the past and modeling. The former is only relevant for the near future, the second collapses as the number of parameters grows. To explain the epistemological limits of the social sciences, Jensen says that there are four essential factors that make it qualitatively more difficult to simulate society than matter: human heterogeneity, the lack of any kind of stability, the numerous relationships to consider both in time and space, and the way people respond to having their activity modeled.

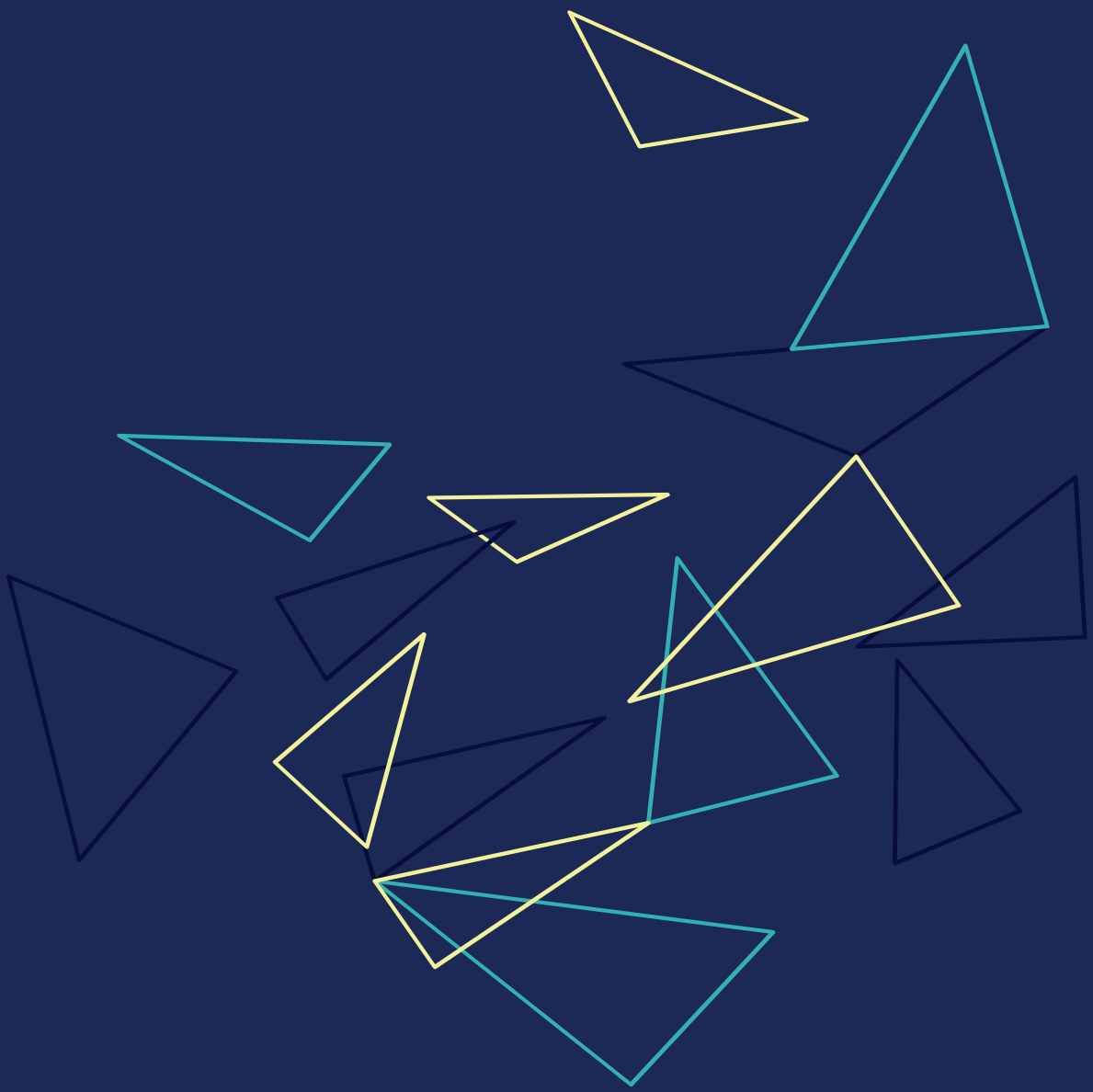
The element to remember here is that the fictional representation of models, which has of course proven to be effective in the physical or natural sciences, is incapable of predicting social behavior. More broadly, modelling and social engineering, despite their rationalist dressing, play a part in the political conception of societies: one that believes that society can be externally modeled and simplified. This ideology of reifying human relations assumes that only action from the outside can "change society" and that the creativity and pluralism of the people involved is not enough. In reality, it is a depoliticized, even dehumanized, view of social relations.

16. Newton: "And thus Nature will be very conformable to her self and very simple, performing all the great Motions of the heavenly Bodies by the Attraction of Gravity which intercedes those Bodies, and almost all the small ones of their Particles by some other attractive and repelling Powers which intercede the Particles." *Opticks* 4th ed., London, William Innys, 1730, p. 372.

Einstein: "...that is to say, the theory, of every natural process, including life," *Einstein's Essays in Science,* Mineola, NY, Dover Publications, 2009, p. 3.

17. Jensen, P., *Pourquoi la société ne se laisse pas mettre en équations*; Editions du Seuil, March 2018.

18. For example, the relation between growth and employment may seem intuitive, but over the long term the correlation proves to be uncertain, hindering reliable forecasting and planning.



Big Data Factory for Big Decisions

The call to data prayer

Claire SOMERVILLE

lecturer, international affairs executive director of the gender centre, Graduate Institute Geneva

"We need data"- comes the call to nearly every pre-decision making process in policy, programmes, politics and processes be it in public panels or internal meetings- the requests for more data keep on coming. And even if some data are available, the call comes "*we need **more** data*".

To what extent is the echoing ode to data a new phenomenon? Or is the song a response to the multiplying sources of new data purportedly at our fingertips- the raw material of our social, economic and political lives.

Either way, decisions-makers- those with executive and also corporate powers- appear to have become paralysed in making (public) decisions without recourse to "the data". We have moved from the mantra of evidence-based policy, through policy-based evidence to data-driven evidence and policy. The United Nations Sustainable Development Goals 2030 Agenda (SDGs) is a base in point on the call to data. "Trusted, accurate data is key to make sure we move forward on the right track" cried the head of the UN Global Working Group on Big Data at 4th International Conference on Big Data (UNDESA). With 17 goals, 169 targets and 230 indicators – the institutional framework to measure success only on the basis of what can only be "big data" is a mammoth enumeration endeavour that are supposed to hold those in power accountable for their investments. The monitoring and evaluation of the SDGs over a 15-year time span may become the world largest Big Data experiment.

But have data always been so important?

To reflect on the pre-digital revolution when data were something mainly scientists generated with often little connection to the worlds of action and decision--making, brings us to question what we define as data and, furthermore, wonder about how our worlds are reproduced in these new data streams. Are our decisions and actions any better now with all these so--called big data at hand?

And what is this data to which the calls to prayer are invoked? What are data and where is this mystical reservoir of magical essence that will make all human decisions better? Why do leaders and decision-makers feel they need data to move forward?

In an era of unprecedented mistrust, fake news and weakened governance decision-makers appear to lack confidence to act and take responsibility without first letting the data orb whisper it's guidance not unlike the oracles of the Azande much studied in early anthropology (Evans-Pritchard, 1937).

The following chapter takes these questions as a landscape against which to respond to some rather more prosaic questions of our time. What are data these days? What biases might such data hold and reproduce over time?

What is Data?

What scientists inside the academy call data often differs from the kind of things referred to as data, especially "big data", by those outside academia. Data is, perhaps traditionally, something that is generally purposefully collected to respond to a fundamental research question. It follows method and methodology, speaks to ontological and epistemology claims and is "typically obtained by scientific work and used for reference, analysis and calculation" (OED). Raw data, as Dourish and Cruz (2018) recently commented, is an oxymoran – and further still, data, they argue, must be narrated to give shape and meaning. To present meta or big data, as is so often the case, as the "raw material" of human life is to eschew the complexities of data capture and collection, sampling, representativeness, bias, and what van Dijck calls "datafication" (2014). These "big" concepts underpin the science behind data and without which the production of any data risks spurious claims.

Most of that which is described as "big data" is one way or another closely tied with the digitalization of everyday processes and activities. Where once a telephone directory was just that: a place to identify a name, address and landline telephone number in order to contact another person; now it is a searchable, codeable, analysable source of understanding and mapping of geographies, genealogies, migrations, health, education, social and economic status, religion, voting patterns… and the list goes on as such directories are augmented with other big data sources such as google maps, ISP analytics, electricity and water meters, bus timetables and more. The possibilities are exponential and the augmentation apparently seamless. The more we digitalize the deeper and broader our data sets become. So much so that modern data scientists no longer analyse but "mine" their data seeking out that gold nugget on which they can claim their fortunes. The data-rush is here to stay- and so we must become more cogent of the status, use and mining of such sources and the limitation as well as

opportunities they afford us.

What marks out these "new big" data is what McFarland calls "found data" (2015). They arise from observational sources; what I call the outputs of the digitalization of everyday life. They are not purposely sought under scientific rules of design and rigor. They are just "there"; the by-product of other efficiency saving processes of everyday life. Returning to the telephone example: itemized billing of mobile calls together with mobile 4G tracking means it is possible to document the everyday movements and contacts of ordinary people: it is possible – but for what purpose? Whereas, somewhat in contrast to these digital outputs, the academic scientist commences with a research question(s) and considers what data would need to be collected to respond to the question (and then goes out and collects it with a suitable sampling methods and size and appropriate method in a specified period of time), these new sets of mass data are collected for very different purposes with no guiding research question. Their purpose is, for example, to ensure a user pays the correct bill in full knowledge of their usage. Before such invoices were itemized, households simply trusted their service provider to request the correct amount! Trust it would seem, and its corollary, confidence, are two of the unanticipated and undesirable side effects of the rise of these big data.

Whilst not purposeful or designed with the sorts of methodological rigor that scientists inside the academy have spent several hundred years developing, these new forms of big data, these "found data" or by-products of 21st century life, are thought to hold some hidden, previously unknown, possibly un-named aspect of humanity. We can see our lives in ways never known before; we examine streams of sensor data emanating from devices we wear, use, engage and interact with- not even always knowingly.

Data produced as by-products of the ever-growing digital world can and should be subjected to the same or similar processes of rigor and method as those collected purposefully as the raw material of the human sciences. All data, purposeful or by-product, harbour bias – and it is these that I shall discuss by asking whether the sorts of bias that we find in purposeful data are also replicated in by-product data and secondly do are some of the data cleaning and bias reduction strategies employed in the social sciences applicable to big data.

Big Data Bias?

The potential for big data to generate big bias and therefore inaccurate findings is a risk that must be addressed if the intrinsic value of scaled data is to be realized. If we just take a couple of standard methodological concepts from the academic sciences we begin to expose a few of the risks.

Take sampling. The natural and social sciences have developed multiple techniques applicable to quantitative, qualitative and mixed method data collection to ensure that sampling bias are limited in data sets- from calculating P-value significance in hypothesis testing in statistical data to immersive thematic saturation in thick ethnographic data. Academics have a toolbox filled with techniques to design and selection samples and ensure the connected concept of representativeness is fully implicated in all analyses processes. Sampling in the big data sciences is a nascent field and scholars report the "random" is the approach most typically adopted by big data science miners (Rojas et al 2017). Kandel et al, in an interview study of

data scientists, found that data miners were actually wary of using data sampling for fear of (ironically) bias it could introduce to their analysis – and furthermore is contrary to the goal of big data which seeks to use as much data as possible and run experiments at scale. And so we face an inherent contradiction as the new, and yet unproven, big data sciences and scientists try to define the old rules of science.

Let's take another well-founded source of bias in data: missing data. Anyone who has ever engaged in process of "cleaning" a dataset is well-aware the challenge of missing data. Be it as simple as a date of birth or inconclusive blood result on a patient record, to lost files, human error, data entry errors and technology glitches- even the most seemingly straight forward of data counts – the number of people currently living on the planet- are subject to unquantifiable levels of missing data (Wardrop et al., 2018). Missing data can take many forms in big digital data sets and are caused by both human and technological forces including infrastructure outages, update errors, trolling, spam bots even human non-compliance with data capture instruments such as wearable technologies. Sarah Pink and colleagues have begun conceptualizing the gaps in digital data as "broken data" (Pink et al 2018) – and include additional processes that can affect the quality of big data such as decay, repair, re-making and growth. Drawing on ethnographic detailing, Pink begins to demystify the multiple ways in which big data are constituted in all its fragmentations, incompleteness and contingent relations and entanglements with humans as producers but also technology and software as collectors. The materiality of these data cannot live in isolation and do not necessarily have objectively reliable predictive capacities. Missing, decaying and entangled data are intrinsic biases that need to be addressed in the new data paradigm and we will have to re-think our data cleaning processes- as despite its sanitized connotations, big data may well be creating a bigger laundry basket.

Final Thoughts

Taking just two of the key methodological concepts employed in science to understand bias in ordinary data shows us that these are also challenges for big data. The daily work of sampling, selection, accounting for missing and incorrect data are as relevant to big data as to small. Why then is there just now such a call to data as the source of all answers, the reservoir of solutions that have until now escaped our sight but now become visible through digitalization? To a large extent the "call to data" that opened this chapter, and one I hear so often among decision-makers, points to a deeper problem – one of trust, responsibility and belief. The compelling quest to make data-based decisions is in part constructed around the institutional scaffolding of Big Data thinking. The promises heralded by the big data firms peddling and mining away with supercomputer capacities are chipping away at the rather more human capabilities of intelligence and decision-making *sans* data. It is not data itself that renders and speaks but human analyses of data that typically try to simplify complexity and generate "readable, portable and tractable" (Latour 1987) insights. The data brokerage and mining of companies like Cambridge Analytica serve as salient reminders of the fragility and also ethics of using the by-products of digitalization as a source for action.

Since so few of the big data we refer are purposefully sought as part of a methodological design we are left with troubling situations where nearly anything counts as data, especially if it can be quantified or used

to first create and then operationalize algorithms; these mystical formulas known only to the data miners who profess their credibility. Big data comes with the allure of sanitized, objective mirroring of the world and implies as Jasanoff suggests "a panoptic viewpoint from which the entire diversity of human experience can be seen, catalogued, aggregated, and mined so that the narratives derived from the data speak as for themselves, compelling reasonable people to action" (2017). Yet taking just two examples of data bias make visible the cracks in big data "science". Should we be compelled to act in response to the facts produced by big data mining? Even if yes, we should bear in mind a healthy anthropological spin of all things factful: "Anyone can produce a new fact; the thing is to produce a new idea" said Evans-Pritchard of the claims of the witches, oracles and magicians of the Azande.

EVANS-PRITCHARD, E. E. 1937. *Witchcraft, oracles, and magic among the Azande.*

DOURISH, P., E. CRUZ. 2018. Datafication and data fiction: narrating data and narrating with data- Big Data and Society 1:10

JASANOFF, S. 2017. Virtual, visible, and actionable: Data Assemblage and the sightlines of justice. Big Data and Society 1:15

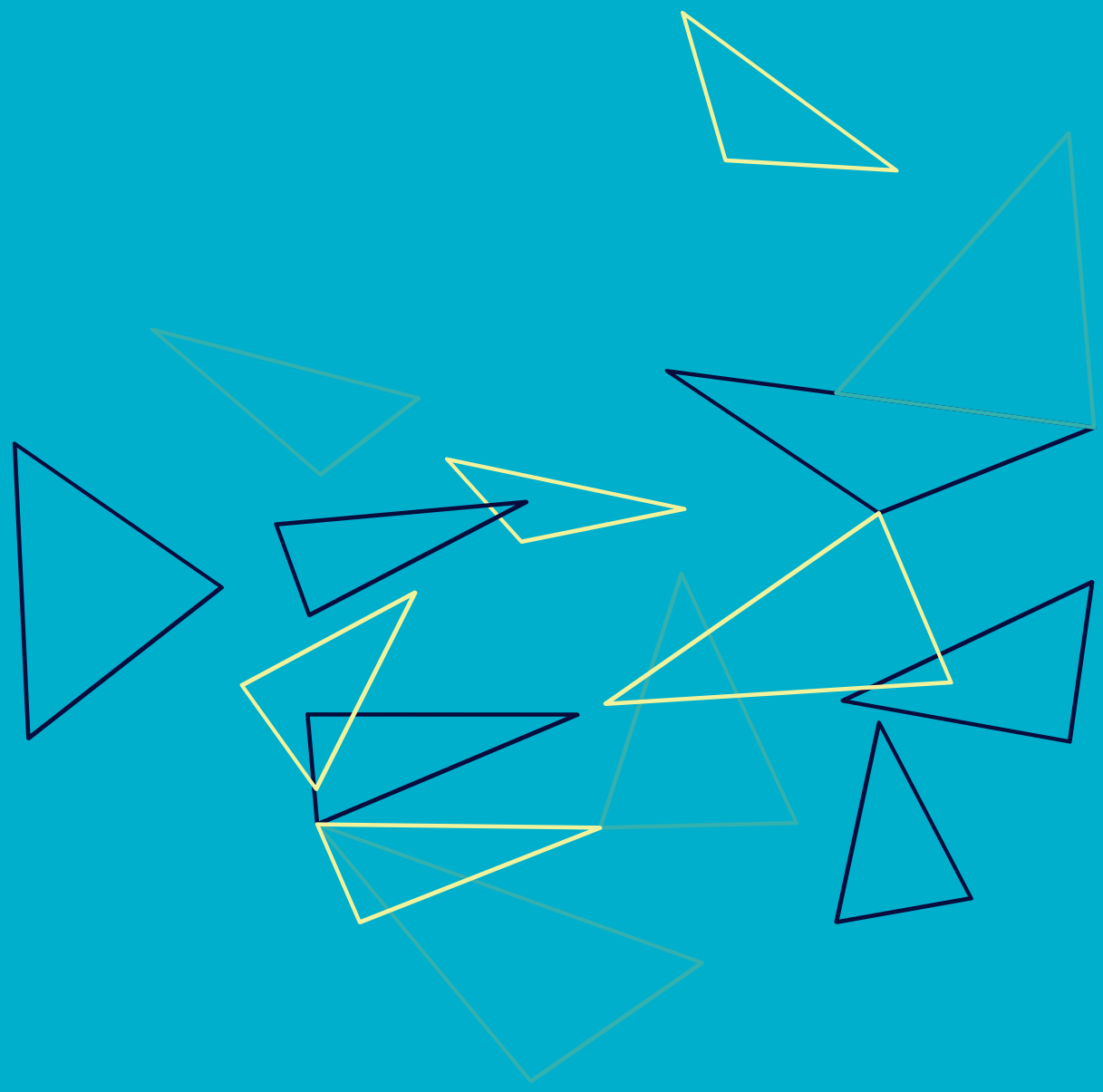
PINK, S., RUCKENSTEIN, M., WILLIM, R., M. DUQUE. 2018. Broken data: conceptualizing data in an emerging world. Big Data and Society 1:18

McFARLAND, D., R. McFARLAN. 2015 Big Data and the danger of being precisely inaccurate. Big Data and Society 1:4

UNDESA https://www.un.org/development/desa/en/news/nocat-uncategorized/big-data-for-sdgs.html (access 01/10/2018)

van DIJCK, J. 2014. Datafication, datism and dataveillance: Big Data between scientific paradigm and ideology. Surveillance and Society, 12 (2): 197-208

WARDROP, N. A., JOCHEM, W. C., BIRD, T. J., CHAMBERLAIN, H. R., CLARKE, D., KERR, D., BENGTSSON, L., JURAN, S., SEAMAN, V. & TATEM, A. J. 2018. Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences.*



The ethical issues of brain-machine interfaces

Laure TABOUY, *PhD Neuroscience*
Bernardas VERBICKAS op, *Vilnius University*

In this chapter, we seek to explore the possibilities offered by the confluence of neuroscience and informatics as a lens for examining the intersection of future relationships between people, data, and machines. A common feature of science fiction, recent progress renders plausible the idea of directly controlling digital devices through thoughts. By the same token, "brainjacking" of implants is becoming a possibility. Given this impending merger of digital and brain data, the authors of this paper seek to take a pedagogical approach and present the state of the art of various existing technologies. We will begin with the significant advances in brain imaging and the first conclusive studies of more invasive methods involving the brain. Afterwards, we will look at these developments from a legal standpoint, considering broad categories of rights of persons, especially in the case of experiments on brain-machine interface procedures.

Neurotechnologies, neuroscience, neuroethics.

The mission of neuroscience is the study how the human brain works. Through it, basic anthropological realities are seen in a new light. The profile of neuroscience has grown in recent years as vigorous debates have arisen over individual freedom, screening and treatments for brain diseases, controlling and modifying behavior, and enhancing individual performance. These discoveries and new methods that neuroscientists, aided by new technologies and therapies, are using in their research make it possible to better understand physiological and pathological pathways as well as figure out what conditions might lead to disease.

Society's fascination with neuroscience and everything it touches, even from afar, is the origin of a number of international projects, such as the Human Brain Project and the Blue Brain Project, which is actually a repository of images and data relating to neural circuits.

There are many questions raised by neuroscience and neuroimaging. In addition, unprecedented questions have arisen about interpreting brain images as well as about screening and treating neurological and psychiatric diseases – which may actually alter brain function – at all stages of human life. It is undeniable that we are experiencing a "neuro-revolution". The latest idea to come out is that knowing how the brain works would serve as the basis for a greater, more complete understanding of human nature, which would trigger discussions, discourse, and fantasies and be characterized by the emergence of interdisciplinary fields. These fields, known in French as *neuro-disciplines*, include neuroeconomics, neurophilosophy, neuroethics, neuro-law, neuromarketing, and neuro-education.

According to Bernard Baertschi, scientific advances and the applications they make possible raise fundamental ethical questions. These concerns are even more pivotal because they involve an organ that, for many, symbolizes humans themselves. The better we

know the brain, the better we know ourselves; to do something to it is to do something to our identity. How far can we and should we go? Beyond strictly ethical problems, the area of neuroethics extends to basic philosophical questions that are are given new, deeper life by neuroscience: the nature of human beings, the body-soul relationship, free will, and personal identity[1].

Baertschi believes that neuroethics builds on and responds to the study of these issues, which in turn leads us to reflect on the role of emotions in our moral decision making, how responsibility and personal freedom stand the test of cerebral determinism, observing mental states through neuroimaging (reviving an old dream), mind reading, and the promise of neuropharmaceuticals to enhance human ability.

Neuroethics is a nascent field that not all academics can agree on. It straddles two worlds, neuroscience and philosophy, and belongs to the realm of bioethics. By subjecting neuroscience to philosophical rigor and, conversely, studying how its advances force us to rethink our moral understanding, Baertschi thinks solidly reasoned arguments would head off both rejection of the principle and naive enthusiasm.

It is now clear that neuroscientific progress and methods of brain mapping and analysis must be considered in both the pathological and non-pathological context. However, the field's increased societal profile provokes harsh criticism that accentuates some of the limits and constraints of imaging and therapeutic technology. More importantly, it highlights unavoidable ethical questions that are serious but vital. Asking and responding to these questions ensures that this technology will be used more appropriately in the future. The current guiding ethical principles are not enough for this field even though the basic groundwork has already been laid by the field of medical ethics[2].

In France, the idea to enshrine neuroethical principles in law came in 2009 in the runup to the revised bioethics law of 2011. It pays particular attention to banning neurobiological discrimination in labor, public health, and insurance legislation.

Hervé Chneiweiss, research director at the French National Center for Scientific Research (CNRS), noted in a 2013 essay on neuroscience and ethics that one of the essential questions of the human mystery is to understand the brain, which allows us to think and communicates the world to us while letting us communicate with the world. Scientific research, neuroscientific advances, new imaging technology, and discoveries lead scientists and society to take a stand on bioethical issues. Data protection, early prediction of diseases in healthy people, and criminal responsibility are just examples of the ethical concerns raised by these studies and discoveries.

As Hervé Chneiweiss says, we must identify troublesome areas where these amazing scientific achievements and discovered treasures might open us up to dangers that need to be spotted and prevented. The dialogue between neuroscience and law is delicate, showing us the importance of ethics within this area of excellence. According to Chneiweiss, there is a tension between the necessary developments to analyze human brain function and the difficulty we have in explaining that scientific knowledge trades in probabi-

1. Baertschi, B. *La neuroéthique : Ce que les neurosciences font à nos conceptions morales*. Paris: La Découverte, 2009.

2. Including: the Helsinki Declaration of ethical principles first established in 1964 for medical research on human subjects.

The Belmont Report, a 1979 declaration created by the American National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research

The Asilomar statement on precautionary principles, published this year and including signatures by business leaders and scientists.

lities, not certainties. The issue today is to know how brain mapping, by letting us see certain cerebral functions, opens new doors that go beyond science and medicine to affect all of society. This includes practices and uses but also imaging of individuals and, when analyzing the neural pathways of decision making, the very concept of freedom of thought and thus renewed discussions of personal responsibility.

The pivotal question today is the individual significance of data collected by any scientific means. A major political and economic section of our society seeks to find in it the bases of individual determinism in behavior.

There are two types of ethics in research: the ethics of basic research – the focus of research procedures themselves – and the ethics of the consequences and applications of neurotechnology.

One is freer to move about in territory that has been charted. Filling the gaps requires reflection and charting all of the different possible paths neuroethics can take. For that, there are areas of concerns that are important to consider: protecting privacy and consent, individual personal identities, enhancement concepts, and correcting existing biases. Questions of transforming personality and perception after operations or implants of chips, electrodes, or medical devices are crucial.

Brain-machine interfaces also pose a number of ethical dilemmas. Scientists use them already in many areas, such as to help paraplegics regain control of their limbs or to steer drones. Prosthetic robots controlled by the brain that let paraplegics be more independent are almost ready for use in everyday environments. The rediscovered ability to grasp a cup of coffee, put away a credit card, or sign a document with a pen increases independence and self-determination for severely paralyzed people. However, introducing devices controlled by brain-machine interfaces into everyday environments that could increase the ability of the able-bodied to interact with digital devices raises a number of ethical and social challenges in the following areas: autonomy, responsibility and accountability, data security and confidentiality, and managing end-users' expectations when there is promise of medical advances in a certain field[3].

The fear of "losing oneself" in a machine may seem overstated at the moment given current BMI capabilities. However, with the field's staggering growth over the last few decades, we should expect rapid developments in what these technologies can offer.

Who is in charge? Who will take responsibility for the consequences of technologies conceived, designed, parameterized, and tested in research laboratories? How do we know if there are ethical issues and what urgent, essential ethical quandaries are raised by using and applying neurotechnology, including brain-machine interfaces? What is their responsibility in these innovations? How can we spot potential dangers, questions, and problems if we do not actually know where to look? The problems of responsibility are already here.

Cerebral imaging techniques and brain-machine interfaces, state of the art

Neurotechnology is at the intersection of human cognition and computer science and involves one of the most fragile and sensitive parts of what we consider

3. A study published in *Science* in June 2017, (Clausen J et al, 2017; *Help, hope, and hype: Ethical dimensions of neuroprosthetics*) agrees that it is time to look into the ethical issues that will undoubtedly arise when mind-controlled computers become possible.

the self. Deeply integrating interfaces into people's nervous systems poses complex ethical questions linked to the definition of a human person as a legal subject and our inalienable dignity. This is all the more true when those techniques affect the brain, with proven effects on personality as has been shown in many cases of deep neurostimulation in Parkinson's patients. Do scientific advances, especially in healthcare and medicine, contribute systematically to achieving the aims of care as defined by our ethics and values? Is it possible that they contradict and subvert these values? Despite these concerns, medicine's techno-scientific approach also risks seeing care as merely a package of pharmacological or technical responses to scientific problems.

These techniques are revolutionizing our approach to the brain and how it works, allowing us to think very differently about neuroscience and pathology.

Technology is ushering in a new dimension, one of a system of direct links between the brain and a machine that makes it possible to carry out a task while bypassing nerves and muscles. The devices let people use their thoughts to control computers, machines, prostheses, or other automated systems. The first human trials started back in the mid-90s. Brain activity is recorded using invasive or non-invasive techniques.

What is neuroimaging?

Brain imaging studies are not easy to design, conduct, or especially interpret. For one thing, the images obtained are measurements, recordings of a specific signal at a given time in a given patient, and thus in a specific context. Great caution must therefore be taken when interpreting the images and signals. A recording taken in two different contexts may have two functionally different meanings. One legitimate and very common debate is over the credibility given to the color-coded pictures of brain activity that serve to fascinate and, by tapping into people's emotions, give the illusion that the images are rich in scientific content. In his book *Neuroscepticisme*, Denis Forest[4] opines on the matter of the trust placed in neuroscience. Through imaging technology, neuroscientists humanize the brain while considering, the person more or less as a whole, which in turn leads to real confusion. Neuroimaging's leaps and bounds of recent years create the illusion of seeing brain activity in real time. Denis Forest, however, urges caution when it comes to hasty conclusions drawn by studies and publicized by the media. These advances are made by developing high-performing tools and machines that improve the granularity and reliability of data.

The following are examples of non-invasive methods:

1) Transcranial magnetic stimulation

Easy access to this method, used in treatment of depression, merits considering its possible unethical applications. It could potentially be used for non-medical purposes, such as to control people. Hopes of neuro--enhancement, already offered online by some businesses could run too high.

2) Electroencephalograms

This is currently a preferred type of brain-machine interface. It is non-invasive and used to restore function after a brain injury. Current research is looking into restoring or enhancing brain function with neurofeedback, in which a subject receives information on their own brain activity in real time. The aim is to use

4. Denis Forest, *Neuroscepticisme*, Ithaque, October 2014.

this method to alter pathological brain activity in the long term, thus changing or restoring the related behavior or injury-related loss of function.

3)PET scan

The pairing of positron emission tomography with an X-ray scanner, known as a PET scan, is a medical nuclear imaging method that enables 3D measurements of metabolic and molecular activity of a specific organ with molecular precision. The key principle of brain PET is to assess the link between neuronal activity in a particular region and the radioactivity of a radionuclide tracer in that same region. PET scans are on the rise in neuroscientific research laboratories and in hospitals. The advantage to this method is that it can specifically target a particular molecule, neurotransmitter, or neuroreceptor and thus produce a very accurate picture of neuronal connection activity. It is an excellent diagnostic tool for a variety of neurodegenerative and neuropsychiatric diseases.

4) Functional magnetic resonance imaging

Magnetic resonance imaging, or MRI, made its medical debut in 1980, where it has ever since allowed doctors to see the structure and internal organization of a living subject's brain and get information about the brain's anatomy that had previously only been possible to get post mortem. It was not until 1991 that scientists discovered that this method could be used to follow blood movement. This groundbreaking discovery in the history of imaging made it possible to observe the brain's reactions in real time and see its activity by following oxygen consumption in its regions. It was a game changer for studying the brain in the areas of science, philosophy, and of course medicine. With increasing advances in speed, resolution, and data analysis, MRI is a credible research tool in neuroscience for examining neuronal and cerebral mechanisms.

The following are examples of invasive methods:

5) Brain-machine interfaces

These devices are used to measure brain activity. There are a variety of techniques in use, the most invasive of which is electrophysiology, or implanted sensors. Currently, the only type of research being done with this method is in medicine, with the aim of eventually restoring function lost due to injury or disease. Studies on brain-machine-interface technology require high levels of ethical vigilance, such as when it comes to using it for enhancing brain performance in a healthy person by recognizing a particular mental state (fatigue, sleep, concentration, etc.).

Non-medical uses may be for civil, military, or commercial purposes, which is why it is necessary to always be on the lookout for potential ethical and privacy violations. It is thus vital to contemplate what constitutes just use of such interfaces.

Current brain-machine interface technology is generally oriented toward therapeutic results, such as helping people with brain and spinal injuries. Already, users can complete relatively simple motor tasks such as moving a cursor on a computer or controlling a motorized wheelchair. Moreover, researchers can already rudimentally interpret a person's neural activity using functional magnetic resonance imaging – such as determining whether someone is thinking of a person as opposed to a car.

These advances could revolutionize treatment for a number of afflictions, from brain injuries and paralysis to epilepsy and schizophrenia, and change the human experience for the better.

A neuroscientist paralyzed by amyotrophic lateral sc-

lerosis (ALS) used a BMI to lead his laboratory, write grant applications, and send emails. During that time, researchers at Duke University in Durham, North Carolina showed that three trained monkeys implanted with electrodes could act as a "brain net" to move a robotic arm together. These techniques can work from thousands of miles away if the signal is transmitted wirelessly via the internet.

6) Deep brain stimulation

This is a neurosurgical technique based on electrically controlling neural circuits via an electrode implanted in a region of the brain. It is an invasive method that involves surgically implanting electrodes into the brain that are connected to a pulse-generator box implanted under the skin that provides a weak electrical current to certain deep regions of the brain. It is used to treat movement disorders and psychiatric diseases with debilitating symptoms that do not respond to treatment (Parkinson's, OCD, tremors, dystonia, etc.). The risks to consider are numerous and can come from the surgery itself, the implanted materials, or even undesirable side effects such as apathy or impulsivity. This promising technology is being fine-tuned to decrease side effects and increase expected effectiveness while maintaining strict ethical standards.

7) Optogenetics

This scientific technique combining optics and genetics is widely used in neuroscience research laboratories. In it, scientists use the genetically targeted expression of light-sensitive proteins known as opsins to take optical control of cells. Its simplicity enables researchers to better understand how the brain works in different pathological and non-pathological conditions. However, although it is in use in neuroscientific research facilities, it is unlikely to be used to treat human brain diseases because it is very invasive, requiring the use of genetic engineering techniques along with an optical fiber implant in the patient's brain to control the modified neurons. The only studies underway in the laboratory are ones investigating eyesight restoration. Optogenetics is considered a type of brain-machine interface because it involves a device being implanted in the brain.

8) Sensory neuroprosthetics

This method started with the success of cochlear implants for the hearing impaired. Ocular implants (retinal prostheses) are electronic implants that restore a field of vision to the visually impaired (such as cases of age-related macular degeneration or pigmentary retinopathy). Currently, sensory neuroprosthetics have no non-medical applications; their use is still exclusively clinical. On the other hand, using this type of method to enhance eyesight poses obvious ethical concerns.

Brain-machine interfaces and legal issues

Since these technological innovations deal directly with the human brain, there are a number of legal questions that classical medical ethics has not yet answered. How do you guarantee free and informed consent? How do you safeguard against discrimination? How do you protect personal data? Who is responsible for damages? Do brain-machine interfaces undermine the integrity of the human race?

One main issue is that of defining identity. A person is unique from conception and their identity is slowly built along with their autobiographical narrative, a

phenomenon known as "descriptive identity". A medical device implanted in the human brain might affect that narrative identity. Such changes in identity could be the key indicators of one of the potentially problematic effects of BMIs. These changes represent a "reality constraint" and are limiting not due to the person's level of desire but because of the implications of deep brain stimulation, or DBS.

Deep brain stimulation raises questions on the legal evaluation of a person's status during treatment. This relates to the legal responsibility of the medical device manufacturer or clinical investigator and to the principle of *restitutio in integrum*[5] of damages. Do changes to and continuity of personality constitute legitimate, legally protected interests? Which criteria in DBS treatment define those personality changes that are considered to be negative or harmful? Human experimentation has created so many scandals that it influences any judgement to be made on the ethics of it.

To understand the pivotal ethical point in scientific research, particular attention must be paid to this aspect: Where are the imbalances between the scientists and their subjects and how do you address it to avoid any ethical conflict?

Abuse can also occur when scientists – consciously or unconsciously – prioritize the interests of their research over those of the participants in the therapeutic trials.

Different countries and peoples with different religious, ethnic, and socio-economic backgrounds will have different perspectives. As such, governments must create their own deliberative bodies to ensure a mediated, open debate involving representatives from all parts of society as well as to figure out how to translate these guiding principles into policy, including specific laws and regulations.

As neurotechnology and businesses develop and governments and other parties work to ensure that citizens have new skills, their identity (physical and mental integrity) and ability to act (freedom of choice) must be protected as basic human rights. One possibility might be to add clauses protecting the rights all people vis-a-vis neurotechnology in international treaties.

On the other hand, international laws and declarations are merely agreements between states. Even the Universal Declaration of Human Rights is not legally binding. It may be necessary to create an international convention together with the United Nations to define banned actions as concerns neurotechnology and artificial intelligence similar to the bans listed in the 2010 International Convention for the Protection of All Persons from Enforced Disappearance.

These declarations must also protect people's right to be informed about the possible cognitive and emotional effects of neurotechnology. At the moment, consent forms generally only focus on the physical risks of the surgery, not the possible effects on humor, personality, or sense of self.

It is hard to predict which technologies will negatively impact human life, and so any line that is drawn is inevitably going to be blurred. It is thus vital that guiding principles be set at the international and national level to place limits on emerging neurotechnologies and to define the contexts in which they can be used. This is the case for human genetic modification, the use of human embryonic stem cells, and cloning.

Some cultures place a higher value on privacy and individuality than others. As a result, regulatory decisions must be made in a specific cultural context while respecting universal rights and global directives. In addition, outright bans on certain technologies may well push them underground. As such, efforts to establish

5. In law, *restitutio in integrum* means restoration to original condition and is the standard consequence of a nullified contract.

specific laws and regulations must include organized fora for thorough, open debate.

These efforts should take their inspiration from their many predecessors that have established an international consensus and included public opinion in scientific decision making at the national level. A conference held in 1925, in the aftermath of World War I, led to the development and ratification of the Geneva Protocol, a treaty banning the use of chemical and biological weapons. In the same vein, after World War II, the UN Atomic Energy Commission was created to examine peaceful uses of nuclear energy and control the proliferation of nuclear weapons.

Military applications of neurotechnology must be strictly regulated. For obvious reasons, any moratorium must be worldwide and sponsored by a UN-led commission.

Trust or concern?

Brain-controlled implants are no longer the stuff of science fiction, which is why scientists are contemplating how to point out the potential dangers of these brain-machine interfaces.

One day, a brain-machine interface will cause bodily harm to someone. To resolve the issue of responsibility raised by such damages, we will likely need to devise a system that asks human users to approve actions by the machine they are interacting with or to refuse any actions they do not want. Users of brain-machine interfaces could approve or abort a robot's actions with an eye-tracking system. This system would not work as well if the robot itself is defective, but manufacturers and lawmakers are already experienced with issues of product risk. It would therefore require creating a new risk-evaluation regime for brain-machine interface technology.

Another major problem involves privacy. Brain-machine interfaces have the power to reveal a great deal of physiological information without the user's consent. We already put a large amount of our private lives on computers that are vulnerable to hacking and there is currently no reason to believe that BMIs would be less susceptible. It is unlikely that brain data would not be bought and sold in the same way the personal information we share online is. BMI companies need to develop clear ethical rules on how this data will be stored and used.

Technology uses encryption to protect data. What type of encryption do we need to safeguard our brainwaves? Might it be worthwhile to escalate neurosecurity to prevent unauthorized manipulation of neural data, or "brainjacking"? For the moment, there are not many answers to the numerous questions raised by brain-machine interfaces; the research goes on, already unveiling their potential in helping to treat paralysis or even concentration disorders. If we want to avoid a future in which millions of people are brainjacked in a large-scale hacking attack, though, it would behoove us all set the conditions of human-machine symbiosis now.

Protecting biological data recorded by brain-machine interfaces is another cause for concern. Security solutions must include data encryption, anonymized information, and network security. Directives are already in place to protect patient data in clinical studies, but the rules vary by country and do not apply to purely human research in the lab.

According to professor Niels Birbaumer, protecting the neuronal data of people who are totally paralyzed and use a BMI as their sole means of communication is particularly important. The settings for their BMIs depend on the brain's responses to personal questions provided by the family. Strict data protection must be

in place for everyone concerned, including of the personal information asked for in the questionnaires and the neuronal data that ensures the device is working properly.

It will one day be possible to decode people's mental processes and directly manipulate the brain mechanisms that make up their intentions, motions, and decisions. People will even be able to communicate with others simply by thinking. Powerful computer systems linked directly to the brain will make it easier for people to interact with the world, greatly improving their mental and physical abilities. However, this could also exacerbate social inequality and give companies, hackers, governments, or anyone else a new means of exploiting and manipulating people. BMIs could also profoundly alter certain basic human characteristics: the privacy of thoughts, individual agency, and an individual's understanding of themselves as an entity bound to the body.

Some of the world's richest investors are placing their bets on the crossover between neuroscience and AI, investing in the creation of methods that could both "read" human brain activity and "write" neural information to the brain. It is estimated that the for-profit sector expends some 100 million US dollars per year on neurotechnology, and that number is rising rapidly. In the United States, more than $500 million in federal funds have been allocated since 2013 to developing neurotechnology within the US initiative BRAIN alone. The current status is already staggering.

In order for neurotechnology to take off in general consumer markets, the methods have to be non-invasive, low risk, and much less expensive than than current neurosurgical procedures. Still, even today, companies that develop these devices must be held responsible for their products and be guided by certain best practices and ethical norms. We would like to highlight four topics of concern that call for immediate action. Although we are raising these questions within the context of neurotechnology, they equally apply to AI.

Privacy and consent

As mentioned previously, the information you can learn about someone from their personal data trail is extraordinary. Researchers at the Massachusetts Institute of Technology in Cambridge discovered in 2015 that careful analysis of motor behavior revealed by keyboard strikes on personal devices can help in the early detection Parkinson's disease. A 2017 study suggests that mobility model measurements, such as those collected from people carrying smartphones with them throughout the day, could be used to diagnose the initial signs of cognitive difficulties due to Alzheimer's.

Algorithms used for ad targeting, calculating insurance premiums, or matchmaking are made much more powerful by the addition of neuronal information, models of the activity of neurons associated with certain attention states. Internet-linked neuron devices open the possibility for individuals or organizations (hackers, companies, or governmental bodies) to track or even manipulate what happens in a person's mind.

Citizens must have the ability and the right to keep their neuronal activity private. For all neuronal data, the default option must be that it is not shared and it must be under constant protection. Refusing sharing by default would mean that neuronal data would be treated in the same way as organs or tissues in most countries. People can explicitly choose to share neuronal data from any device. This would involve a safe and secure process that would include a consent procedure that clearly states who will use the data, for what purposes, and for how long.

Neuronal data could be used to draw conclusions about people who do not choose to share them. Buying and selling the data and its uses must be strictly regulated and limited in the same way as the option for people to abandon their neural data or to have neuronal activity written directly to their brains in exchange for money. Another privacy protection measure would be to restrict centralized processing of neural data. Using other technologies especially devised to protect personal data would also be useful. Blockchain technology, for example, allows data to be tracked and audited and "smart contracts" provide transparent control over how data is used without the need for a centralized authority. Finally, open data formats and open source code would provide greater transparency on what remains private and what is shared.

Identity and personality

Some people who receive deep brain stimulation via implanted electrodes report a change in their activities and identity. In a 2016 study, a man who used a brain stimulator for seven years to treat depression said he started wondering whether his interactions with others, which to him seemed inappropriate after the fact, were due to the device, his depression, or whether it represented a deeper truth about him.

Neurotechnology has the potential to markedly disrupt a person's sense of identity and action. It has rattled fundamental hypotheses about the nature of personal, legal, and moral responsibility. If automatic learning and devices connected to the brain allow a quicker transition from intention to action – perhaps using a type of "autocomplete" or "autocorrect" function – people might end up acting in ways they will have a hard time claiming as their own. If people can control devices remotely with their thoughts, or if multiple brains are connected together for collaboration, our sense of who we are and where we are acting will be disrupted.

Enhancement

When brain-machine interface technology is used for non-therapeutic purposes, many applications come to mind whose intentions are not to repair but rather to enhance, such as ones that greatly increase endurance or sensory or mental ability. These applications are likely to alter societal norms, raise questions about equal access, and create new types of discrimination.

DARPA[6] and the US Intelligence Advanced Research Projects Activity are in talks on plans to provide soldiers and analysts with enhanced mental abilities, which makes it more difficult to draw the line between repair and enhancement. What will it mean if ocular implants can be used to enhance eyesight to include night vision?

Increased risk of bias

When scientific or technological decisions are made based on a limited number of concepts and systemic, structural, and social norms, the resulting technology may advantage some and disadvantage others. These biases may be entrenched in neural pathways given that it is very difficult to define equality in a mathematically rigorous way.

Industry and academia are already talking about practical measures to counteract prejudice in technology and are coming up with algorithms and other mechanisms to ensure that biases are corrected from the

6. Defense Advanced Research Projects Agency (US Department of Defense)

initial phases of technological development.

A call for responsible BMIs and neuroengineering

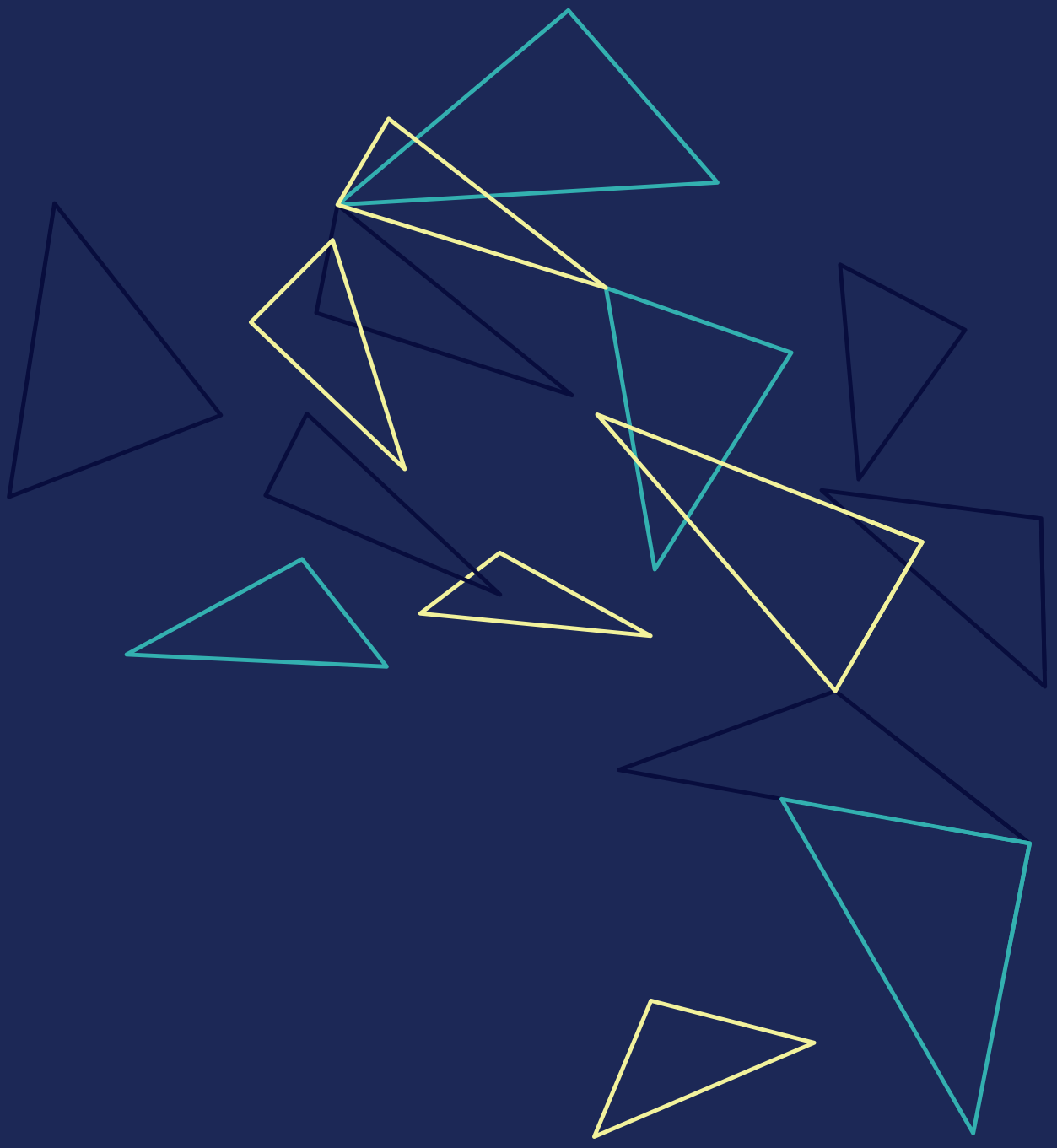
Scientists have already demonstrated the possibility of hacking vital implants such as insulin pumps or cardiac defibrillators. Malicious manipulation of a device could lead to the death of its user. It is also possible to intercept and manipulate biological signals that have been converted to digital ones (bluetooth or wifi). Unfortunately, as the authors of the paper note, there is no known, technically viable solution to this problem.

These essays add to the appeal to industry and academia to assume the responsibility that comes with designing devices and systems that can drive such change. Even now, they can get inspiration from responsible innovation frameworks that have already been developed that also encourage innovators to anticipate, reflect on, and commit to promoting opportunities for socially desirable science and innovation and public-interest initiatives. In business, profit-seeking often wins out over social responsibility. Most of these technologies are created to benefit humanity, but will likely come up against complex ethical dilemmas that they are unprepared for.

Mentalities must be changed by integrating a code of conduct into industry and academia. This involves learning to think more carefully about how advances can be pursued and implementing strategies that are likely to contribute to society constructively rather than breaking it. Neurotechnology has great potential to benefit medicine and society. To reap those benefits, we need to guide their development in a way that respects, protects, and enables the best in humanity.

It is worth noting as well the concerns about higher risks associated with increased radiation doses through repeated scans. An emergence of radiosensitivity has been observed. Another issue must also be mentioned, that of the accuracy and interpretation of images and imaging data. Caution must be exercised when reading into hasty and uncertain conclusions and extrapolations, which are often based on low sample sizes. Every person's brain functions are unique and it is important to consider the human being as a whole.

Just because an image shows a person's brain activity during real or simulated behavior and one moment in his or her life does not mean that this person cannot behave any other way and that other behavior would necessarily be associated with different brain images. The image of a person's brain only shows their activity in that moment. It says nothing of the history that shaped that person as an individual or the brain activity that led to the choice of observed behavior. It says even less of how they will behave months or years later in an environment unbeknownst to the experimenters.



Can social portability for data restore trust?

Lionel MAUREL, *co-founder of La Quadrature du Net*

It has become a common refrain that the internet is having a crisis of confidence, but just what do people mean by that and what does it look like? For example, 2018 will be marked by a long list of scandals involving the social network Facebook[1]. This litany provoked a worldwide backlash, raising serious questions about the risks that centralized platforms pose to the integrity of democracy. In the ensuing months, a number of platforms and social networks were struck by major security breaches, compromising the data of millions of users. This past December, a leak of confidential documents made matters worse by revealing that Mark Zuckerberg's network had struck secret agreements to provide companies like Apple, Microsoft, Yahoo, Amazon, Netflix, and Spotify access to private user data. However, all of these setbacks have also given rise to solutions aimed at recentering the discourse around the notion of the commons being the keystone in human relations, including in the digital world.

1. Lapowsky, Issie. "The 21 (and counting) biggest Facebook scandals of 2018". WIRED, 20 December 2018: https://www.wired.com/story/facebook-scandals-2018/

When the information went public, people were outraged, creating the hashtag #DeleteFacebook to encourage users to close their accounts. After going viral for months, the movement yielded some results: Three million users are said to have quit the platform in Europe. However, this is negligible compared to the 1.4 billion users registered worldwide. Facebook is still growing well enough to continue attracting massive investment from advertisers.

The public's reaction seems to be contradictory. On a collective level, the harm Facebook does seems to be getting more and more obvious. A January 2019 poll shows that internet users trust Facebook less than any other company, far less than Twitter and Amazon[2]. However, usage figures show that, on an individual level, many users find it hard to take the plunge and quit the platform. This could be explained – as it so often is – by the privacy paradox[3]: In absolute terms, people generally place value on protecting their privacy, but they have trouble following through on that, especially when it comes to managing their digital lives.

There may well be another explanation stemming from a lack of confidence on the part of online communities themselves. Leaving a dominant platform is a complicated choice to see through on an individual level because it means the individual has to cut the meaningful, emotional ties that he or she has maintained with other people on that platform. In this situation, nobody wants to be the first person to take the leap or risk being the only one to cut themselves off. We find ourselves faced with what game theory calls the "prisoner's dilemma"[4]: a situation where individuals have to make choices in a context of uncertainty that pushes people to find a solution that may make sense on an individual level, but is suboptimal on the collective level.

Thus is can be said that, on this type of platform, each person may be theoretically free to leave at any time, but communities are no less "prisoners unto themselves"; the thread of social relationships becomes a net that entraps users. The ability of platforms to use the power of social ties against their users equates to a formidable enforcement power that regulations should offer suitable protection against. This, however, is not currently the case. Although the law stipulates that everyone's personal data should be protected *individually,* it still has great difficulty legislating the same data on the *collective l*evel[5]. At the moment, our social relationships have no type of legal recognition: Even in legal texts dedicated to personal data, there is no notion that would allow social connections to be considered as such.

One possibility to fill this gap would be to implement a type of "social portability" for personal data to al-

2. Boule, Marie. "Facebook obtient le pire score pour la confiance des utilisateurs, selon un sondage". Vice, 3 January 2019: https://www.vice.com/fr/article/gy7ea3/facebook-obtient-le-pire-score-pour-la-confiance-des-utilisateurs-selon-un-sondage

3. Laugée, Françoise. "Notre intimité en ligne ou le 'privacy paradox'". Revue européenne des médias et du numérique, July 2018: https://la-rem.com/2018/07/notre-intimite-en-ligne-ou-le-privacy-paradox/

4. Poundstone, William. *The Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb.* Doubleday, 1992.

5. Maurel, Lionel. "Comment sortir du paradigme individualiste en matière de données personnelles ?" S.I.Lex, 19 July 2014 : https://scinfolex.com/2014/07/19/comment-sortir-du-paradigme-individualiste-en-matiere-de-donnees-personnelles/

low user communities to act collectively to break the vise-like grip of the dominant platforms and move to other spaces with more respect for privacy.

Protecting privacy, a collective issue

As sociologist Antonio Casilli rather provocatively puts it, there is nothing more collective than a piece of personal data[6]. A statement like this might seem counterintuitive at first glance since personal data tends to refer to that which is private, intimate, confidential, and thus individual. This is also how the law sees personal data, given that it is defined in legal texts as "any information relating to an identified or identifiable natural person"[7]. Personal data is thus ruled on – and protected – solely in terms of its ability to identify an isolated individual.

And yet, this "individualist approach" fails to include other aspects inherent to personal data, such as data that defines our social relationships. Indeed, our private lives are part and parcel of our social lives, involving our romantic partners, friends, family, colleagues, fellow club members, etc. As such, "personal" data is also always – to varying degrees – "social" data. Indeed, this is what dictates how digital platforms collect our personal data and extract value from it. Going back to the example of Facebook, it is notable that the company is actually less interested in information relating to a particular individual than in being able to figure out their location on the "social graph".

"Social graph" is an expression that the Palo Alto-based firm uses to refer to the way they record relationships between users. Soon after it launched, Facebook realized that this human map was the real source of value to be harnessed through selling targeted ads. Mark Zuckerberg referred to this explicitly in 2007: "If you take all the people and all their friends in the world, that constructs a social graph...Facebook [doesn't own] the social graph, there just *is* a social graph of the world. What we try to do is model that and map it out. We're not creating new connections....We're trying to map [the world] out exactly."[8]

As a result of its underlying individualist presuppositions, the current law treats personal data on a granular level, but not in aggregate, and it is precisely from here that the big platforms derive their power. The Cambridge Analytica scandal showed that all it takes is for a company to convince 270 000 users to take a quiz to be able to vacuum up the data of 87 million Americans by tapping into to Facebook's social graph. Thus a series of individual actions had a massive collective effect while simultaneously revealing the weaknesses in the legal perception of the very nature of data.

Even today, some people go further and view privacy

6. Casilli, Antonio and Tubaro, Paola. "Notre vie privée, un concept négociable". *Le Monde,* 24 January 2018: https://www.lemonde.fr/idees/article/2018/01/24/notre-vie-privee-un-concept-negociable_5246070_3232.html

7. Article 4.1 of the General Data Protection Regulation: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679#d1e1489-1-1

8. Transcribed and condensed from an interview by Michael Arrington with Mark Zuckerberg at Techcrunch40. https://www.youtube.com/watch?v=vkGke4UWDCk
Zuckerberg's comments start at 1:15.

as a "commons". This is how Jon Evans described it in a recent article on Techcrunch written in response to the umpteenth Facebook scandal[9] in which the platform offered teenagers 20 dollars in exchange for installing an invasive application that collected their personal data. Jon Evans points out that although the consequences for the individual were minimal, they were potentially much greater on the collective level:

Ok, maybe you think rootcerting a teenager is sketchy — but if an adult chooses to sell their privacy, isn't that entirely their own business?
The answer is: no, actually, not necessarily; not if there are enough of them; not if the commodification of privacy begins to affect us all.

[...]while individually, our privacy may usually be mostly meaningless, collectively, it is a critically important commons. Anything that eats away at our individual privacy, especially at scale, is a risk to that commons.

The question is therefore the following: How can we legally protect the commons that comprises our privacy and social connections?

Abandoning the choice between public and individual data portability

In late 2018, the *New York Times* revealed that Facebook had shared its social graph data with certain companies like Apple, Microsoft, and Amazon through secret agreements[10]. Creating an ecosystem of applications that reuse its data was one way for the company to make itself indispensable. The published documents also showed, however, that it could deliberately choose to deny a competitor access to this resource and thus stymie its development. This was the case for Vine, for example, an application specialized in video that Facebook eventually considered to be too dangerous a rival.

It is thus clear that Facebook's social graph plays the role of what is known in competition law as an "essential facility", which the French Court of Cassation defines as a facility or infrastructure which is necessary for reaching customers and/or enabling competitors to carry on their business[11]. Generally, public authorities are not supposed to let these types of resources fall into the hands of one company. Thanks to its social graph, Facebook finds itself in a dominant position, able to control access according to its own interests and not that of the general public.

To remedy this situation, some solutions have been proposed that would enable "public data portability", or rights granted to the public authority to force platforms to open and share their data. Essayist Evgueny Morozov claims that states should even give the whole of their population's data a status of public property,

9. Jon EVANS. "Privacy is a commons", TechCrunch. 10 February 2019: https://techcrunch.com/2019/02/10/privacy-is-a-commons/

10. Dance, Gabriel J.X., LaForgia, Michael, and Confessore, Nicholas. "As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants". *New York Times*, 18 December 2018: https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html

11. Verdier, Henri. "La donnée comme infrastructure essentielle". *Rapport Etalab,* 2016-2017. https://www.etalab.gouv.fr/wp-content/uploads/2018/04/RapportAGD_2016-2017_web.pdf

thus allowing it to be licensed to private companies for a fee[12]. Other proposals are less "collectivist", saying that the state should be able to declare certain information linked to strategic sectors (security, health, transportation, energy, etc.) as data of general interest, obligating private companies to return them or share them. Similar ideas can be found in places like the Villani report on artificial intelligence[13] which insists on the need to create various "data commons". According to the report, public authorities need to seek out new modes of production, collaboration, and governance of data by creating these "data commons". These commons would incentivize economic stakeholders to mutually share data. The state's role would be as a trusted third party. In some cases, the public authority could mandate that some data of general interest be shared.

The biggest problem with these proposals advocating for "public portability" of data is that they require placing trust in the state as a mediator and regulatory entity. Such trust is on the wane, however, as states – even "democratic" ones – enact security policies that rely on intrusive technology. When it comes to mass surveillance, we have known ever since the revelations by Edward Snowden that states and the major platforms have been in collusion. These conditions make it treacherous to give the state the power to requisition personal data that would enhance its powers beyond that of supervision.

The right to individual personal data portability, one of the innovations introduced in 2018 by GDPR (General Data Protection Regulation) is a more classical approach. According to the French National Commission on Informatics and Liberty (CNIL), this means that people have the ability to retrieve some of their data in an open, machine-readable format. They can also store or transmit that data easily between information systems to reuse it for personal purposes[14]. This right is sometimes portrayed as a "counterweight" in the hands of consumers to create competition in the digital sphere. It allows them to actually recover their data and transfer it to another service that they deem more useful.

The problem here is that people do not make much use of this right despite it being enshrined in law. As we have seen, people maintain their social connections through digital interactions on platforms and the very force of these relationships dissuades people from invoking their right to personal data portability. As with many other aspects of GDPR, this right was conceived of by considering the "granular" level of personal data, but not the aggregate level. Platforms themselves have understood very well that this right poses little threat, to the point that companies like Google, Twitter, Microsoft, and Facebook have forged an alliance as part of the Data Transfer Project[15] to implement an open source tool that encourages people

12. Maurel, Lionel. "Evgueny Morozov et le domaine public des données personnelles". S.I.Lex, 29 October 2017: https://scinfolex.com/2017/10/29/evgeny-morozov-et-le-domaine-public-des-donnees-personnelles/

13. Cédric VILLANI. "Donner du sens à l'intelligence artificielle : pour une stratégie nationale et européenne." February 2018 : http://www.enseignementsup-recherche.gouv.fr/cid128577/rapport-de-cedric-villani-donner-un-sens-a-l-intelligence-artificielle-ia.html

14. CNIL. "Le droit à la portabilité en questions." 22 May 2017: https://www.cnil.fr/fr/le-droit-la-portabilite-en-questions

15. https://datatransferproject.dev/

to exercise their right to personal data portability...
To get out of this predicament and reconcile public and individual data portability, we must redefine the concept and imagine a "social" data portability.

Establishing "social portability" for personal data

A proposal of this ilk was put forward by the nonprofit La Quadrature du Net in late 2018[16]. The initial idea was to leave behind the misleading notion that platforms are mere "passive hosts" and recognize the "enforcement power" that they impose on their users. This power manifests itself in the fact that platforms are not "neutral" about the content they spread since they rank it using algorithms. Their power is also evident, though, in their ability to take our social connections and use them against us. In their proposal, the group states that the tech giants' enforcement power could be the criterion that restricts their new status. This "power" arises when users of a platform cannot leave it without suffering "negative consequences", which allows the platform to impose its own rules. In the example, these negative consequences are the loss of human connections made on the platform.

The purpose of a law is to rebalance power relationships by making them legal relationships. Platforms have an enforcement power on the very threads of our social relationships, and so the law needs to impose protections in the form of interoperability. Again according to La Quadrature du Net, in reality, we have no choice but to continue using the tech giants in order to not lose the connections we have made on them. This is something that could be corrected if the tech giants became interoperable with other services, if they let us keep talking to our "Facebook friends" without having to stay registered on Facebook ourselves.

La Quadrature du Net also says that, technically, "interoperability" would come via "communication standards", or multiple services using a common language to communicate with each other. For example, ActivityPub is a standard for "decentralized social networks" that gives us a concrete reason to hope for the rise of the decentralized web. Also, using these standards would be a way of making the GDPR's "right to portability" effective. Without inter-platform interoperability, it has failed to prove its utility.

In addition, the nonprofit organization says we could quit a tech giant, such as Twitter, in favor of another service, such as Mamot.fr, or the decentralized microblogging service Mastodon that La Quadrature du Net offers. With the new service, users can continue to send and receive messages from people who remain on the tech giant (Twitter) without having to cut ties.
Nowadays, there are decentralized services that provide technically convincing alternatives to the major, centralized platforms and do not exploit their users' personal data. This is the case for Mastodon, an equivalent of Twitter or Facebook, as well as for Peertube, an equivalent of YouTube[17]. The thing holding

16. Messaud, Arthur. "Régulations des contenus : quelles obligations pour les géants du web ?" *La Quadrature du Net,* 9 October 2018: https://www.laquadrature.net/2018/10/09/regulation-des-contenus-quelles-obligations-pour-les-geants-du-web/

17. "Peertube : Le logiciel libre est une alternative crédible à l'hyperpuissance des GAFA". *La Tribune,* 15 October 2018 : https://www.latribune.fr/technos-medias/peertube-le-logiciel-libre-est-une-alternative-credible-a-l-hyperpuissance-des-gafa-793324.html

users back is not the lack of alternatives but rather the challenge of cutting themselves off from their relationship networks.

This is exactly why we need to create not just individual portability for personal data but "social portability". Each individual would still get to choose whether to move from one platform to another or to a decentralized service such as Mastodon. However, this choice would be made much easier since it would no longer involve severing connections made with other users. What matters is not so much that personal data is portable, but rather what happens to our connections on the social graph. By making interoperability mandatory, a platform like Facebook would no longer have the "captive" audience it does today.

This would bring us a type of "collective personal data portability", but without state intervention or entities charged with representing the will of various groups. It also avoids having to forego individual consent while still allowing individuals to operate in a new framework that is more conducive to personal and collective emancipation. Reworking the law would recognize for the first time the importance of protecting social connections as something more than personal data.

ETHICS & TECH Report 2019

Part II
NEW SOCIAL CONTRACT, NEW GOVERNANCE

2018 will go down as the year of the communal and salutary wake-up call to the definitive entrance of digital issues in the field of politics. Between the revelations linked to the American election, parliamentary hearings in the US and Europe, increasingly large and frequent security breaches, financial assaults on more and more states, or even the enactment of the European General Data Protection Regulation (GDPR), the last several months have been punctuated by a long litany of affairs highlighting, in our view, a new era of public awareness.

As such, beyond a certain initial shock which may affect public trust in these technical devices, it is also necessary to measure the positive consequences of developing a social consensus on the necessity of strongly repoliticizing the matter of technology's "entrance into society".

We feel that this movement of "technical democratization" which emerges from different contexts and recent experiences must be highlighted as a positive issue that should be encouraged at all levels.
The following contributions, without straying from a critical analysis of the ideological processes of depoliticization at work in the expansion of digital technologies symbolized by the concept of "State-platform", return to recent controversies linked to the confrontation between the political and technological fields. Thus, and beyond the common assumptions about Chinese initiatives in the sector, we reflect upon the reality of the now-famous Social Credit System being rolled out and the moral panic it engenders.

While the overall discourse insists on the so-called innovative nature of the situation, we must nevertheless return to some fundamentals of the political aspect, such as governance, sovereignty, collective mobilization, or even representativeness as the various chosen authors here are committed to it.

Of course, while the massive entrance of digital questions into the political field is a fortunate evolution, solutions must still continue being developed collectively, step by step, in order to face the challenge of maintaining a social contract which is still anchored locally but also, in part, increasingly de-territorialized.

With the participation of:

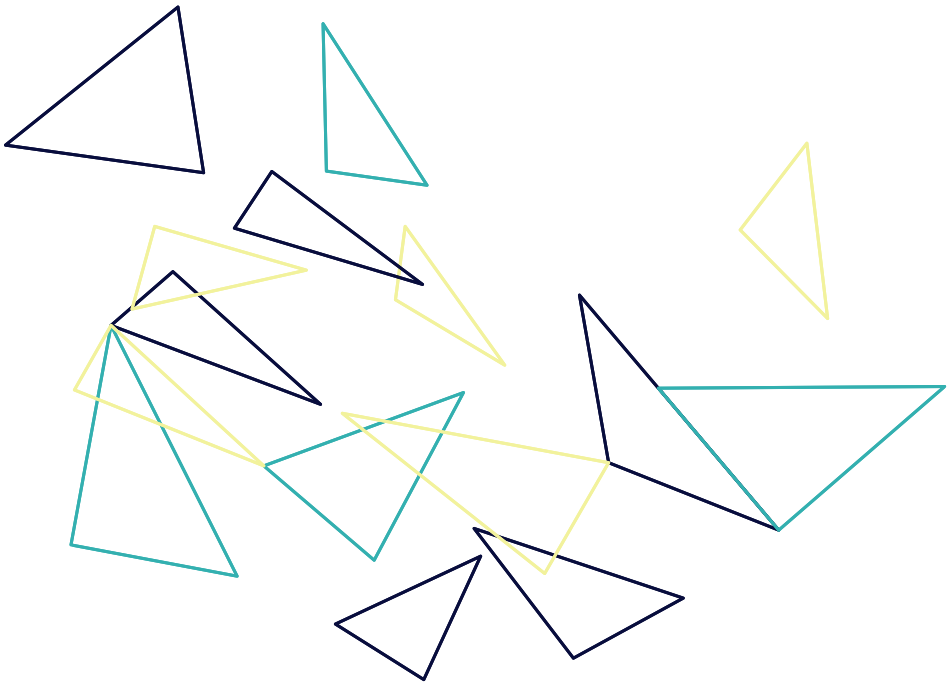
Marylaure **BLOCH**

Fabrice **EPELBOIN**

Pierre **GUEYDIER**

Helen **MARGETTS**

Adrian **PABST**



CONTENTS

Who governs digital technology?

Dr. Pierre GUEYDIER, Director of Studies - OPTIC
Dr. Adrian PABST, Dean of Political Science faculty - Kent University

What powers govern digital technologies? Does the extension of digital technology transform contemporary systems of government, including the administration of rights of "property" in the cybernetic world and its relationship to the common Good? Governance of digital technology raises fundamental questions concerning representation: what interests (individual or general) and what ideas or ideologies are represented in the political sphere and its institutions? In turn, representation affects the creation of laws and policies of regulation. In short, governance is not limited to standards, regulations, and infrastructures, but extends to matters of power, society, and justice.

Back to the Estonian case

In order to better understand the operational principles at work in digital governance, one case must be considered: Estonia. This country of 1.3 million inhabitants wished to build its post-Soviet identity and its soft power on the thundering affirmation of a "100% digital" state. Apart from an obvious argument about geopolitical singularization tied to its geographical and historical situation, this laboratory can, after several years' experience, establish a useful case without claiming to account completely for the stakes of governance in the digital age and its methods.

The Estonian case is a unique, full-blown mode of assimilation of digital technology into the public and political spheres alongside three other grand models: (1) China and its "dystopian and totalitarian" approach; (2) Russia, which would militarize data as much to internal ends as to external ones; (3) Western democracies which attempt to conform the digital revolution to liberal principals[1].

Naturally we must try to go beyond a too-simplistic analysis in order to grasp the distinctiveness of Estonia and its possible limits. The key of the Estonian context rests on the notion of "mutual accountability"[2]. The logic of "by design" transparency, embedded in political discourse and technical infrastructures, is promoted to the rank of national value as the software of the balance of powers. Not classically between the executive, legislative, and judiciary, but in the sense of reciprocal control of the State and the citizens. The asymmetry of relationships between citizen and state and political history do not, however, make a case for a natural self-limitation of any form of power. In effect, the building of everyone's trust in such de-

1. Andrew Keen, *The Internet is not the Answer,* Atlantic Books, 2014.

2. Geoffroy Berson, "e-Estonia: the ultimate digital democracy?", *Medium*, 24 Sept. 2018.

vices has a collective cost: the sacrifice of a large part of the concept of privacy, even though it is essential in democratic history.

Estonian digital transparency reiterates and therefore involves a gradual desacralization of the private sphere, contrary to what Benjamin Constant (1767-1830) had theorized as "liberty of the Moderns". Here we must quote the liberal thinker who denounced the illusion that limiting powers by increasing their number is effective: "The authority which issues from the general will is not legitimate merely by virtue of this[...]. Sovereignty exists only in a limited and relative way. The jurisdiction of this sovereignty stops where independent, individual existence begins. If society crosses this boundary, it becomes as guilty of tyranny as the despot whose only claim to office is the murderous sword. The legitimacy of government depends on its purpose as well as upon its source."[3]

The prevailing fatalism which declares the struggle against the capture of private data to be lost before it begins inexorably causes formerly "sacred" privacy to slide toward a post-privacy era borne not only by political initiatives like the Estonian case but also by the agenda clearly undertaken[4] by the adjutants of digital technology. Going back to Estonia, we also have to consider historical and geographical aspects in order to understand why Estonians develop a trust that may seem naïve in their political and public institutions. Through their recent history following the fall of the Soviet Union, the Estonian people believe in the power of transformation of these institutions which benefit from a high capital of sympathy, probably also connected to the size of the country, which engenders a certain mechanical proximity to the citizens, elected officials, and the administration. The relationship to the concept of totalitarianism, significant in the environment of this part of Europe, absolutely does not have the same collective translation as Germany, for example, or other Eastern countries for whom it is inconceivable that such a blank check should be accorded to a state's administration.

The Estonian authorities defend their digital policy with three arguments. First, in an Eastern Europe with a lagging economy, the "technology leap" has acted as a narration for the collective effort of progress. Secondly, the very low population density has brought the government to vast digital infrastructure plans. Finally – and not without paradox, considering the risk of cyber-attacks[5] – the digitalization and "dematerialization" of the State were defended as sovereign will and political independence vis-à-vis Russia.
Estonia, ultimately, is a case of clearly accepted depoliticization. The displayed objective is thus to create an "invisible State" whose prerogatives would be reduced to those of a provider of efficient public services. The criterion of managerial efficiency thereby favors cronyism over citizenship in the relationship with the people. Does the Estonian experience of "Democracy as a Service" prefigure the future of digital governance?

3. Benjamin Constant, *Principles of Politics Applicable to All Governments,* May 1815.

4. Eric Schmidt, former CEO of Google, affirmed nearly 10 years ago already, on CNBC in 2009, *"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."*

5. Indeed, one of the most tremendous cyber-attacks ever conducted on a national level targeted Estonia in 2007 amid tension with Russia. As a result, NATO installed its cyber-defense center in Tallinn in 2013.

Governance by numbers

At the heart of governance lies the trust of the governed toward the governors and cooperation between the representatives and the people. And yet we are currently having a crisis of popular trust and public cooperation. Western democracies have, in the past, faced periodic crises characterized by a breakdown of trust in politicians and public institutions, low popular participation in the political process, and a profound skepticism of the capacity of the democratic system to solve urgent problems or serve the long-term interests of an entire country. But the current crisis seems qualitatively different: (1) the mutual distrust between citizens and the political class in the contemporary context which philosopher Pierre Manent summarizes as "populist demagogy and the fanaticism of the center"[6]; (2) the function of the digital world is not immediately compatible with the values of the system of representative government. These rest on the idea of government by the people while the advent of digital technology carries with it the idea of "governance by numbers"[7]. Such a concept of governance will ultimately erase physical and cultural borders, submit the nation state and the welfare state to the world market and dismantle the protective rules that govern nature, labor, and currency.

This evolution brings back into question the very idea of national and popular sovereignty. The representative model is also hybrid, mixing elements of democracy, aristocracy, and monarchy. Popular sovereignty is limited not only by the rule of law but also by the prerogatives of the head of state: "the politician is never the double nor the spokesperson of the elector, but he governs in anticipation of the day when the public will render its judgement"[8]. Today citizens' loss of confidence in their representatives undermines the authority of the elected and, consequently, casts doubt on the legitimacy of the laws they vote on, including the authority of public institutions responsible for managing regulation policy.

In addition, the social contract on which the representative system is founded is brought into question by an evolution even more fundamental than the crisis of confidence – the very logic of the representative democracy, which combines equality in the eyes of the law with freedom of thought and expression. Thanks to this double principle, citizens participate in the governance of the political arena, and as the governed they may criticize the governors. However, civic participation and criticism allow the democratic system to function and improve by correcting its own deviations, particularly the concentration of power and wealth that is characteristic of contemporary democracies. Yet digital technology reinforces the extension and the intensity of the globalization process, which is to say that the forces of the "Marché total, peuplé de particules contractantes n'ayant entre elles des relations que fondées sur le calcul d'intérêt. Ce calcul, sous l'égide duquel on contracte, tend ainsi à occuper la place jadis dévolue à la Loi comme référence normative[9]". Inasmuch as numbers replace the law as the

6. Pierre Manent, "Populist Demagogy and the Fanaticism of the Center", originally published in *Le peuple existe-t-il ?* under the direction de Michel Wieviorka, Sciences Humaines, 2012, p. 275–86.

7. Alain Supiot, *Governance by Numbers,* originally published by Fayard, 2015.

8. Bernard Manin, *The Principles of Representative Government,* Cambridge University Press, 1997.

9. Supiot, *Governance by Numbers, op. cit.,* p. 15.

instrument of governance, the law yields to the software program and the system of rules yields to the process of keeping the software functioning.

In the face of this evolution, representation by numbers, which irrigates rights and laws, is further and further from the state of the real world. This gap between representation and reality weakens the credibility of the ethical framework of democracy. Thus, equality in the eyes of the law can hide not only certain people's social-status privileges but also, and especially, differences regarding home ownership – notably private property linked to personal data. Indeed, the extension of the world market leads to a transformation of the very nature of political power – a shift from the ideal of government by the people to the idea of an impersonal government which increasingly takes the form of governance by numbers, consecrated by the notion of "platformization of the state". Such power promotes the calculation of special interests at the expense of the common Good, which is to say the triumph of individual or collective usefulness in interpersonal solidarity. Henceforth competition between individuals considers economic calculation as being more fundamental than the pursuit of a just social order – this is the risk of a return of social Darwinism fed by the abuse of a monopoly of technological platforms.

Governance of infrastructure, governance by infrastructure

The particularity of the internet as an instrument of power resides in the hybridization of multiple materials which give it a veritable Leviathan-like nature faithful to the famous Thomas Hobbes allegory. Indeed, these modes of polymorphic power transcend the limits and the historical players of international law, whether it concern states or international treaties. Governance distributed between the design of technical protocols, the policies of private and commercial operators, and administrative bodies of network operation have ushered in a new ecosystem of power linking technical infrastructures, laws, and multiple stakeholders of all sizes, all on a global scale.

These games of power by proxy in the technical infrastructure have brought about a veritable "infrastructural turn[10]". As such, we can count four interwoven dimensions of internet governance by infrastructure[11]. First, the way the internet is governed is clearly distinct from the way in which it is used. Second, the spectrum needed to analyze internet governance must not be limited to information technology and issues of power linked to software but must extend to the material borders and their tangible effects in such areas as microprocessors or the management of the terrestrial spectrum. Third, the attention given to institutions framing the internet infrastructure (ICANN…) must not hide the impact of design techniques, the political agendas of private companies, local legislation, or

10. Francesca Musiani, Derrick L. Cogburn, Laura DeNardis, Nanette S. Levinson Eds., *The Turn to Infrastructure in Internet Governance,* Palgrave Macmillan US, 2016.

11. DeNardis, Laura (2014). *The Global War for Internet Governance.* New Haven: Yale University Press.

international law. Finally, beyond technophile rhetoric, strategies of access restriction, control, or even censure must be emphasized. As a result, we can outline six functions of the internet governance ecosystem:

(i) Administration of critical resources (management of domain names, for example);
(ii) Establishment of standards and technical protocols (TCP/IP, HTTP);
(iii) Coordination of access and interconnections (between the main undersea and terrestrial cables)
(iv) Cybersecurity policies;
(v) Policies concerning private intermediaries (access providers, hosts, platforms);
(vi) Legal architecture and the management of intellectual property rights

Contrary to popular belief, the internet is an extremely material hybrid device comprised of an invisible arrangement of layers that compose the network of networks. The methods of operation of this reticular arrangement, its performance, and its distribution constitute the heart of the issues of governance.
This distributed, performative, and "invisible" characteristic of internet governance questions the classic concepts, methods, and disciplines of political analysis such as Law or Political Science. The very nature of the powers deployed in this way is disconcerting on account of the complexity of their imbrications, their deterritorialization, and their amplitude. Eventually, the concept of sovereignty, an essential notion since the treaty of Westphalia in 1648, finds itself frontally examined.

Toward a "soft sovereignty"

The notion of sovereignty is the fruit of complex historical constructions. While it fundamentally resides in a "natural" coupling with the notion of territory, the historiography of the sovereignty-territory couple also reveals a strong hybridity – that it concerns the post-colonial period, and more recently the post-cold-war period, and the advent of a globalization inexorably eroding the authority of the states as "sovereign subjects" of international law.
With breaches ushering in the modern period, the foundation of power, progressively defined by general will and popular agreement (social contract), examines the link between sovereignty and territory, its entry into space, and its new weaknesses.

Since their emergence, the authoritative, perhaps dictatorial powers have perfectly analyzed the risks and potentialities of the internet, like China, which combines its leverage effect on economic growth with heightened state control. Democracies, for their part, are caught in the middle between defense of personal liberties and necessary limitation of their extensiveness, for which the modern sovereign state must be both *de jure* and *de facto* guardian, simultaneously. Yet, multiple examples attest to the obvious loss of effectiveness and sovereignty, particularly concentrated in the incapacity to levy taxes equitably, ensure the security of infrastructures, and command respect for the enforcement of the law.

This *de facto* state, the fruit of a long devolution of state prerogatives, is reaching the limit of social acceptance. 2019 will probably be the year of an attempt at an explicit recovery of sovereignty by western democracies on the internet. Beginning in 2015 and in the particular context of anti-terrorist struggle, the long state of emergency implemented by France for nearly two years ultimately enabled provisions of exception

regarding internet control to be integrated into common law[12]. The implementation of the General Data Protection Regulation (GDPR) in May 2018, American Senate hearings with the directors and principal actors of digital technology, or even the first record fines imposed upon those same actors by the European Union for abuse of power illustrate the changing demeanor of state powers.

Nevertheless, this reversal of states' behavior, beyond the media hype, could reveal itself to be a formidable trap. On the contrary, in an operational plan the restoration of a form of state sovereignty might reveal the weakness of western democracies in enforcing the law in the face of digital leviathans. This collective threat to the loss of trust as much in state authority as in technology was emphasized publicly by the French President at the Internet Governance Forum at UNESCO in November 2018: *"Our governments, our populations will not tolerate much longer the torrents of hate coming over the Internet from authors protected by anonymity which is now proving problematic. At the end of 2018, we stand at a crossroads. Not only is the Internet under threat, but the Internet itself is starting to be described by some as a threat, especially in democratic societies."*[13]

The establishment of new state regulations could result in rapidly fragmenting the internet into constellations. A number of initiatives give a glimpse of the appearance of alternative protocols to TCP/IP. Several founding figures (Tim Berner Lee, Louis Pouzin) work explicitly on protocols meant to restore an original internet symmetrically protected against excessive commercial concentration and state regulation. The decentralization of future protocols and their multiplications lead *de facto* to alternatives to TCP/IP protocol, thus multiplying the destructive effects on state sovereignty. Naturally aware of this risk, will the future be an operational redefinition of the concept of sovereignty? Just like the expansion of the procedures of soft law and the emergence of a veritable normativity of compliance, discussions of sovereignty negotiated between sovereign states and operators, on topics like hateful comments on platforms, are starting to take shape. The commandants, under double pressure from the states and from a crisis of reputation amongst their users, seek to establish the actual para-legal procedures of "soft law" in order to take part in the regulation of their content. Thus, an "appeal process" has just been put in place by Facebook for cases of dispute over content being blocked by the social network. Is a negotiated "soft sovereignty" emerging to avoid the explosion of a harmful internet as much for the states as for the commandants?

Will the internet still exist in 10 years ? For western democracies, toxic internet policy is reaching critical levels. Not only the issue of security of information systems, but more symbolically by the exposure of the inherent weakness in democratic systems which promotes the absolute individualization of rights, leaves the way open even for its gravediggers. The performance efficiency of state sovereignty responsible

12. Internal Security Code, article L851-3, Created by LAW n° 2015-912 July 24, 2015 - art. 5: *"I.-Under the conditions laid down in Chapter I of Title II of this book and for the sole needs of the prevention of terrorism, it may be imposed on the operators and persons mentioned in Article L. 851-1 the implementation on their automated processing networks intended, according to parameters specified in the authorization, to detect connections likely to reveal a terrorist threat."*

13. https://www.elysee.fr/en/emmanuel-macron/2018/11/12/speech-by-m-emmanuel-macron-president-of-the-republic-at-the-internet-governance-forum

for communal trust finds a formidable opponent in digital networks. Is a takeover possible? At the end of the day, do authoritarian regimes have reason to maintain strong control over their online exchanges? The question comes to mind just from seeing the legislative evolution of democratic states inexorably attracted to a model of restoration of power.

But a return of authority could provoke a proliferation of protocols with a design that involves an autonomy that is effectively quasi-impossible to regulate. The adjutants of digital technology are no longer interested in a Balkanization of the digital sphere. A convergence of interest thus seems to have to structure a sort of balance of power between democratic states and digital adjutants, one needing to hedge against the anti-social effects of the networks while accepting a limitation of their sovereignty, the other protecting their economic position by agreeing to temper their abuse of a dominant position. The worst is perhaps not waiting for such a situation which, while it admittedly erodes the ideal vision of state sovereignty, could enable a form of balance of responsibilities in the digital sphere. This intense inclination toward the broad "platformization" of social relationships – not just economic, but political too – might prevent the explosion of the network of networks which is undeniably a source of aggravated conflicts.

Social networks, mobilizations, and Democracy

Fabrice EPELBOIN, *Teacher at Science Po Paris.*

Social networks are the projection, in a social space orchestrated by algorithms, of society – or rather, societies. It's a global space where what constitutes boundaries is of a linguistic order and is unique to each one of us. If you speak a foreign language, new spaces open for you, and up to now, there's been no need for a passport to gain access.
It's also a space where each person draws, through connections of friendship, "likes", or subscriptions to groups, a map of their own space orchestrated by algorithms specific to each social network – algorithms which all have in common the effect of enclosing each person in their "bubbles", encounters in territory they've already explored, and recommendations from the algorithm.

Naturally, the darkness of our societies in reflected therein, sometimes accentuated, even transformed by powerful mechanisms of protection specific to the digital world. We often cite Dunbar's number, which suggests that we cannot establish human relationships with more than 150 people, as the very example of a constraint of the real world that is shattered in the virtual one.

Regarding **governance and different political regimes**, certain political regimes have succeeded better than others in launching into these digital territories and in taking a role in them vis-à-vis their citizens. China[1], which learned how to keep its digital sovereignty and thereby even keep control of its destiny, is in the process of building a new model of a "Big Brother" society based on the surveillance and continual evaluation of citizens, determining their access to multiple services (public or private, like credit) and freedoms (like the right to move around).

The Philippines learned how to project a dictatorship and its distinctiveness onto Facebook, in order to find a form of sovereignty therein, through the presence of authorities on this social network but also thanks to a network of militants to whom harassment of political opponents and the defense of the present regime are delegated.

The "progressive" western democracies endeavor to rethink social networks and must still imagine how to project their sovereignty[2] into these spaces, which, failing that, remain in very large part under American sovereignty, as is regularly shown by Facebook's methods of censorship, which don't hesitate to judge a Courbet painting as pornographic but are very lax with a racist speech that is protected by the First Amendment of the American Constitution. This sharing of sovereignty between Facebook and western democracies should materialize, in France, with a law

1. See in the following pages the contribution from Marylaure Bloch on the topic of Chinese Social Credit.

2. See the concept of "soft sovereignty" in the previous chapter.

meant to fight against hate, whose legal definition promises heated discussions.

But for those who, within a society, are **opposed to its governance**, social networks also offer an obscurity that can be a form of protection.
We have thus observed, during the Green Revolution in Iran in 2009 and a few years later during the Arab Spring, that the pseudonymity offered by social networks has enabled an opposition to unify, organize, disrupt, even overthrow oppressive governments.
"Obscure" opinions also find refuge in social networks. For those in a society whose ideas are not reflected in the media (due to censorship or self-censorship), social networks offer shelter for discussing and sharing opinions in relative obscurity.

If a government tries to apply a form of censorship in the country it is responsible for, we systematically find content censored on social networks – peer-to-peer relationships (and the algorithms) play the role of content distribution formerly played by the mass media.
Thus, the ideas and content censored in the media take a disproportionately large scope on social networks, and the communities unified by this content have the opportunity to learn the rules specific to this environment and their uses to political ends well before the communities whose opinions are reflected and promoted in traditional media (individual dialectic, methods of dialogue and of meeting people, etc.).
In France, and for a generation, the communities unified by the ideas of Jean-Marie Le Pen, excluded from the media at the end of the 90s, as well as those brought together more recently by Dieudonné, and more broadly the France of "no" concerning the 2005 European Constitution referendum, have found on the web and on social networks a space of free expression where these communities have been able to develop militant practices – where other French political currents are just starting, for the most avant-garde, to prompt their troops to get involved, without any particularly advanced strategy due to the inexperience of their troops.

Let us add to this complexity that the political parties have never learned how to project themselves onto social networks. In these networks, ideas unify people more than programs – which are just a collection of ideas – thus giving birth to discussion spaces which productively unify individuals coming from a wide variety of current policies, sometimes totally opposing ones. As such, the Citizens' Initiative Referendum (RIC), very popular among the "Yellow Vest" protesters, has been discussed on social networks for over ten years, and millions of people have been reached by these discussions, many of them coming from "extreme" parties (left or right), to the point of infusing respective policies into their programs under the pressure of their respective militants.

Finally, the **algorithm** itself has a dark side. Without going so far as to attribute non-financial intentions to Facebook, this single factor plays a huge part in what we perceive today as the "dark side" of social networks. The main purpose of these algorithms is to optimize the platform's revenue and to proceed, through the magic of "artificial intelligence", with a mixture of content served to us on demand at any time. This algorithm continually improves this mix to generate more "engagement", meaning more time spent on the platform, to offer Facebook what a former director of French broadcasting very cleverly called "available brain time". Empirically, today everything shows that the best fuels the Facebook algorithm has found for

generating more engagement are hatred and anger, which can only be attached to a dark side in our western societies, where these feelings are banned or narrowly limited in the public sphere.

Social networks and mobilization?

The first social mobilization of importance that we can unquestionably attach to social networks is the "Green Revolution" in Iran in 2009. The tool of the time was Twitter, mainly used more for raising international public awareness than for coordinating actions.
The subsequent Arab Spring saw Facebook, which in numerous Arab societies was already a substitute for a social space crippled by their ruling regimes, serve as a tool for mobilization and coordination of a willfully leaderless movement. This principal feature was, at that time, an innovation for a hybrid movement[3].
Since then, the social protests born of more or less coordinated and more or less intentional interactions on social networks have multiplied. From the Indignados in Barcelona to Occupy Wall Street in the US by way of Hong Kong's Umbrella Movement or the opposition to Erdogan in 2013, the number of cases to study from the last ten years is rather sizeable.
In the end, we are in the same situation regarding social movements as all organizations facing digital technology. Two methods of transformation await them: "digital transformation", which essentially consists of improving an existing organization and using technology to improve its performance, and "disruption", which consists of reinventing the organization from possibilities offered by digital technology, to compete with an established organization.

In this perspective, the fate and the transformation of a company, a union, a political party, or a democracy are not that different. Some will know how to transform while avoiding disruption, as is the case with China or the Philippines, each one using their own approach; others, particularly those who haven't taken seriously the work involved in transforming to face this new territory that is digital technology, will be immediately disrupted, as was the case for the Tunisian regime.

The Fifth Republic of France is currently facing this type of disruptive phenomenon with the Yellow Vests, perfectly symbolized by what has become, over the course of the protests, a demand they all carry: the Citizens' Initiative referendum, which is as such a proposition of disruption of a presidential regime conceived in its time to provide stability – though lacking popular legitimacy – to the "winner", and which was very important after a Fourth Republic marked by parliamentary instability.

In these movements, social networks have three major roles.

1/ They serve as an alternative to the media, which militants see as a faction of the oppressors – and rightly so, more or less; the relationship to the power of the media in France and in Tunisia are not comparable, but the silence observed by French media regarding police violence during the first two months of

3. That is to say real and virtual: purely virtual forms of social protest such as Anonymous having already demonstrated the possibility of a leaderless movement, it should be noted in passing that the Tunisians had previously imagined the first hybrid leaderless movement, "Takriz", a mix between Anonymous and Black Bloc.

the movement tends to show critical flaws in a media system, whose original mission was to be an integral part of a democratic system.

2/ They also serve to coordinate actions, which can take very simple forms: a meeting on Facebook, which was the first course of action for the Yellow Vests, just as it was at the origin of Tahrir Square in Cairo. They can also take more complex forms and help to organize a protest movement in a much finer way, dividing up roles and planning more articulate actions.

3/ Finally, they help to feed a positive feedback loop, whose role is to provide the movement with its dynamic and its motivation, which is most often achieved through recycling certain content. In Tunisia, during the beginning of the revolution, this positive feedback loop was composed of captured videos of police violence that were uploaded and shared. One incident of police violence gave birth to two protests, which generated two incidents, which led to four protests, and so on. From a certain dynamic, these positive feedback loops transform a protest into a riot, a riot into an insurrection, and an insurrection into a revolution. Ben Ali, who understood the internet and technology very astutely, was careful to block access to the Facebook page that allowed videos to be uploaded, but relays outside the country were able to recover these videos and upload them from abroad.

In France, we observe these same positive feedback loops with equally effective images and videos of police violence as a result of their undergoing a form of censorship in the media. We also observe a spectacular diversion of traditional media content, particularly the virulent interventions of certain editorialists and politicians within reach taking sides against the Yellow Vests, which are recycled *ad infinitum,* further feeding the anger[4].

Algorithms, editors of information

Particularly on Facebook since the algorithm was changed at the beginning of 2018, it's the individual users who have had the most impact on content distribution. It may be recalled that this change in algorithm turned a flow of information previously determined by a user's "page likes" and content published by the user's contacts into a flow of information composed of content discussed in groups whose membership is linked to the geographic proximity of their contacts. This modification to the algorithm changed, overnight, the nature and the origin of the information that comprises our individual Facebook feed. One of the most visible changes to the algorithm was the dramatic decrease in media pages, which had previously done much of the distribution of the content of said media on Facebook. Today, these media can only reach their readers on Facebook effectively by paying a high fee to the platform or by counting on their readers to pass content along to others with whom the media had lost contact long ago. Community management was not their core business.

It may also be recalled that half the population get informed uniquely through Facebook and that more than two-thirds use Facebook as a news source. In practice, Facebook is not a source but a distributor, similar to an 80s television set except that it does not produce content and that its method of distribution is

4. This also explains the dual position of 24-hour news channels, at once one of the most hated media in France and one with a steadily growing audience.

particularly complex, gives ultra-personalized results (hence "information bubbles"), and will remain whatever darkness may occur (because of the use of algorithms based on artificial intelligence, which hardly enables retro-engineering).

We have therefore entered, much more than before, a newsworthy ecosystem very largely dominated by Facebook which is vested with the capacity to editorialize the news, which is to say, assemble a collection of content which users use to give meaning to the world they live in. A role that was attributed to the TV news a generation ago recurred for a certain elite in a grand daily paper like *Le Monde* or *Le Figaro.* This elite have not really changed their habits of news intake[5], whereas the intermediate and popular classes have radically changed their way of getting news. This explains the appearance, on the occasion of the Yellow Vests crisis, of a dramatic hiatus between the world of journalism and the people.

From the point of view of influence, a massive portion of the population has passed into the hands of social networks (and by extension into the hands of their users and algorithms), and generally more into the function of distribution than into the production of content. This discreet paradigm shift was poorly grasped by the media who see in distribution a function with little added value, a prejudice inherited from the time when this distribution was limited to transmission and a sales network. We'll return to this point later on.

Parallel to this shift over the last ten years of distribution and audience (and therefore, influence) toward Facebook, France has seen the rise of a phenomenon of militancy specific to the regime of individual censorship that has been established in France since the 80s.

Whereas in the USA an identity militant advocating the supremacy of the white race would have no problem finding a medium that reflected their opinion and no hindrance in broadcasting their opinion on Facebook or elsewhere, they would quite alternatively be in a country where racism, antisemitism, homophobia, and many other things are censored and the publication of such content can bring about heavy legal sanctions.

The media territory which has developed in such a legal framework puts militants of such causes in front of the challenge of recreating an alternative media ecosystem by adapting its expression and its militant approach, if only to find shelter from the law. It should be noted that far less dismal causes have had themselves excluded from a large part of the media ecosystem without the need to appeal to the law. Such was the case of the opposition to the 2005 European Constitution, whose partisans also gathered almost exclusively on social networks to exchange views and to campaign, for lack of seeing their opinions reflected in the media. Ultimately, since the end of the 90s, a large part of the public opinion, sometimes represented in a deliberately caricatural way, has thus more or less been excluded from the media.

The result is the development of renewed and particularly effective militant practices consisting of establishing themselves as distributers in order to use, to influential ends, content whose authors had ne-

5. Today they read *Le Monde* AND *Le Figaro*, but on line, imagining that this change of medium suffices for reaching modernity.

ver imagined it would be used in this way. The most striking example in France is a blog called "*Français de souche*" ("Native-born French"), related to the identity sphere which attracts a volume of around 5 million visitors monthly, an audience comparable to that of a major daily newspaper. The blog is actually just a press review, made up of articles that mostly come from mainstream media but read by a community assembled around the identity trend (far right). Thus, a sentimental article published in a left-leaning newspaper imploring its readers to show solidarity with migrants would be served to a radically different audience from the initial target in order to galvanize and unify the community in its opposition of the government's migratory policy.

This type of strategy is at work in the propagation of "fake news", and we have seen multiple resurgences of members of the identity movement publishing entirely factual articles from the mainstream press dealing with the "Marrakesh treaty", accompanied simply by a short commentary intended to put the internet user in a state of mind that leads them to conclude that the information is biased and deliberately misleading even before they read it. These strategies, playing on the effect of repulsion, are not only impressively effective but also of astounding economic means. The blog *"Français de souche",* which thus creates such an audience worthy of the largest French daily papers, is maintained by a single person, while its reports represent dozens of journalists.

This type of rival blog, which settles for being composed essentially of press reviews, has been a common militant practice in France for more than ten years, foreshadowing the shift of the influence of content toward its distribution. These practices of "diverting" distribution (which could be seen as "influence theft") are now common on Facebook. The most scathing example is none other than the positive feedback loops which feed the anger of the Yellow Vests and which are mostly composed of content from the media, content which rightfully opposes the movement.

The crucial question of content distribution

As a result of the change in Facebook's algorithm, pages "belonging" to the media, which hitherto ensured distribution of their content, are now, for the most part, in the hands of the users.

We are, however, in a crisis of confidence[6] where politicians, political parties, unions, and journalists can't even capture the trust of 10% of the population. In these conditions, at best the content that journalists produce is questioned and disparaged, and at worst, it serves as ammunition in sterile dialectic battles between political militants. Sterile because these battles are not meant to convince anyone of anything but rather to indicate their inclusion in one camp and their opposition to another.

Only seasoned, experienced militants know how to use this content to an end of conversion, with the goal of convincing or destabilizing the adversary. These disciplined militants, armed with political content deftly accumulated over time, are apt to disrupt the discourse they may find on the social networks. Thus, it is very easy to destabilize a pro-European with a few selected links to quality sources intended to illustrate a choice of: the system of tax avoidance in Eu-

6. Measured at regular intervals over the last ten years by CEVIPOF/ Sciences Po.

rope, the Kafkaesque side of European decision-making mechanisms, the effects of the monetary policy of the European Central Bank, or the power of lobbies in Brussels.

Of course, we find these veterans among the ranks of the extremes, not because their ideas are more solid than those they oppose, but because they're experts in the field, in their own dialectics, and in the communal game that deviates from the traditional political rules of engagement. As such it is striking that the far-right militants show much less aggression on social networks than those of "La République En Marche!", who largely discovered the political use of social networks during the last presidential campaign. The ensuing dialectic battle wholly resembles a confrontation between the foreign legion and a group of scouts.
The impact on democratic pillars

It is important, in order to respond to such a question, to distinguish, especially in the French context, what falls under democracy and what falls under the order of the Republic. From a strictly democratic point of view, there is nothing much to fear from an angry mob who is demanding the establishment of the Citizens' Initiative referendum. This form of democracy, practiced in Switzerland the same as California, has largely shown this. It is also useful to note, in order to dedramatize the possible consequences in France, that our Swiss neighbors, whose democratic and institutional stability we can only emphasize, have, like us, a political arena dominated by a far-right party and not lacking in provocative, even outright batty, political personalities.

What is threated by social networks, in Tunisia yesterday and in France today, is the Republic and, in the case of France, the translation it made in its constitution of democracy, a representative democracy, into a presidential regime.

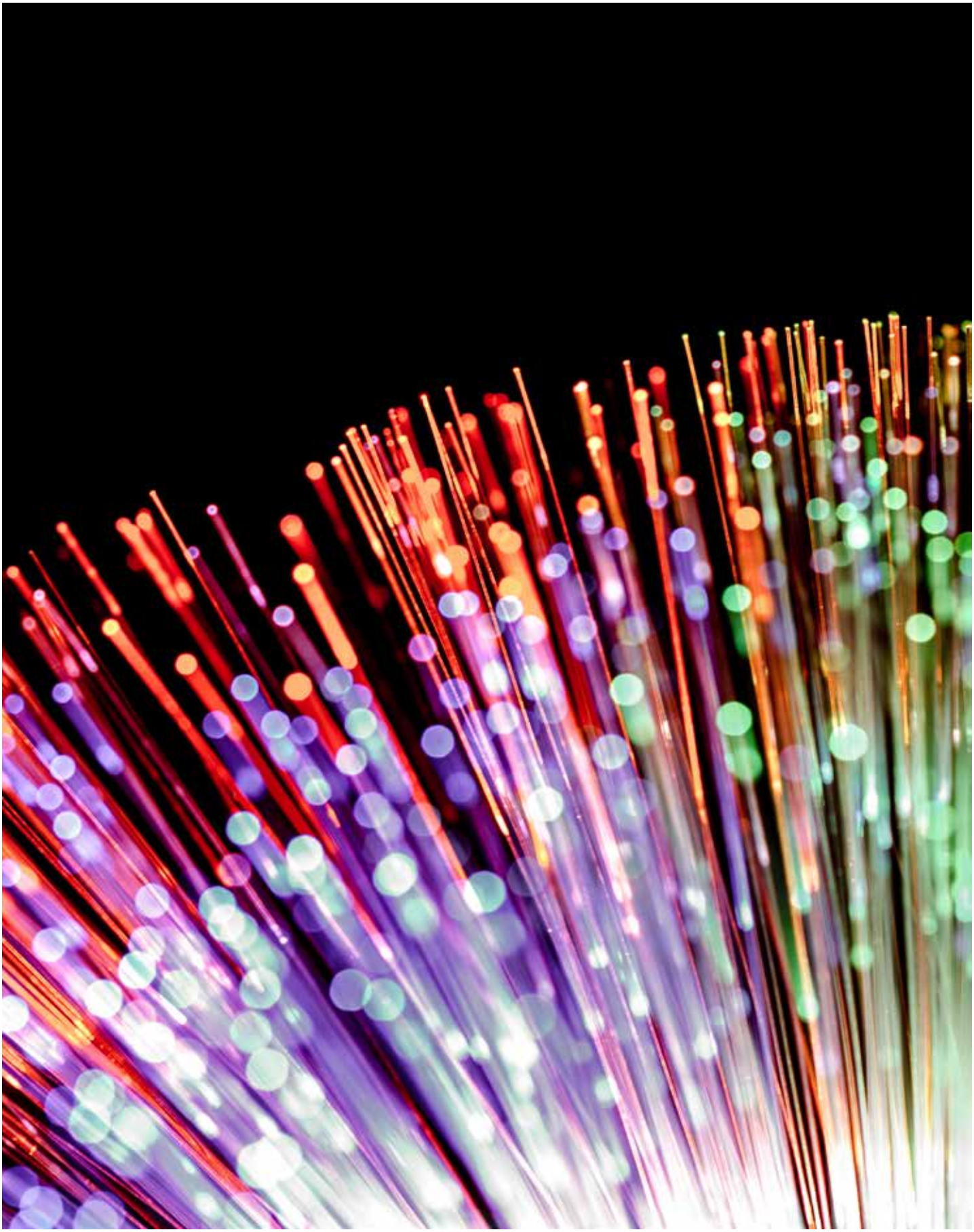
To understand this crisis, widely shared in the French population[7], it is vital to review some fundamental steps of the Fifth Republic, starting with its founding. The French Constitution was written to respond to an institutional crisis: a parliamentary regime, characterized by chronic instability, due to precarious alliances between political parties. It was also written for a man, General De Gaulle.

To compensate for this instability, our constitution gave the one who "came in first" an over-representation in the parliament, in order to secure for the presidential regime a chamber whose political color would be in line with the executive.

While the various co-existences didn't significantly undermine its functionality, the change of a seven-year to a five-year presidential term synchronized the elections of the executive and the parliament, rendering the parliament completely subservient to the executive – which many political scientists consider a problem of democratic order. The arrival of Emmanuel Macron, whose strategy of access to power consisted of circumventing the bipartisanship that constituted the bedrock of our constitution, reinforced the executive even more, rendering the traditional left-right alternation inoperative, generating a feeling of general frustration among the people.

In this context, social networks are merely the outlet

7. 70% of the population think democracy works poorly or very poorly in France / CEVIPOF 2019.



of this democratic frustration, and France is not the first democratic nation to know this type of crisis – far from it. This frustration takes, according to the settings and features of each social network, specific forms. It can translate into the real world in different ways, according to the populations who manage to crystalize this frustration and turn it into a protest movement.

With a feeling of injustice toward the economic system, crystalized with the help of social networks by an urban and majority-student population, we saw the appearance of movements such as Occupy Wall Street in the United States seven short years ago and Nuit Debout in France nearly three years ago.

Today, in France, it's the suburban middle class who crystallize a similar sentiment in a different protest movement whose dynamics are close to the Arab Spring as far as its operation on Facebook is concerned.

None of these social protest movements born on social networks have ever made the slightest hostile claim toward democracy. On the contrary, at the heart of all of these movements is a demand for an increased democracy and a somewhat fundamental re-assessment of the political and economic system. They all have in common a demand that was summed up perfectly in the statement that appeared at the beginning of the Arab Spring: **Democracy and Dignity**.

The Social Credit System in China: Reflection of Our Fears on the Future

Marylaure BLOCH, *doctoral candidate in Contemporary Chinese Studies, University of Geneva*

"Just think: a revolutionary technology, which will disrupt numerous business models, completely transform the economy and society, and what innovation is it going to bring? Trust."

Seeing "Social Credit System (SCS) in China" in the title of this article, some of you probably thought that the epigraph was taken right out of propaganda from the Chinese communist party, praising trust for getting people to swallow the pill of the totalitarian and nightmarish instrument that the country is preparing to put in place. By 2020, every physical person or legal entity will be given a score that will dictate the treatment they will be accorded in society: prove your credentials with a high score or, in the case of a poor evaluation, submit to the pressure of your peers while every computerized system penalizes you, perhaps even forbids you access to certain services as punishment.

What if I told you that in reality, this quotation is just one exaltation among so many others[1] about the promises of the blockchain in society? An ultra-publicized technology which is still in its early stages, the blockchain reflects the importance we give trust in order to develop a healthy society and a healthy economy. Since the Wall came down, realizing that the democratic system alone does not guarantee prosperity, numerous economists and sociologists have emphasized trust as an essential vehicle of development. The discussion around "nudging", popularized by Richard Thaler[2], also includes surprising similarities, to say the least, to the SCS, in that both provisions seek to influence human behavior in a not very coercive way. Without judging those who would be misled, I wonder. About the vast Chinese experimental project, sure, but that goes beyond the scope of this article. I also wonder why attempts to improve society by delegating trust to technology are, in the case of China, laden with negatives it seems, while for the West, it's a long-awaited revolution. Through the prism of thick-skinned Westernism, we distrust the Chinese rhetoric, even if we use similar words on our side.

Put in the spotlight these last two years, the first appearances of the term Social Credit System (**社会信用体系**) actually date from 1998, when the Chinese government was looking into solutions to build trust and, with time, promote the socioeconomic development of the country. At the time, it was difficult to imagine all the potential technology that would see the light of day. The purpose was above all economic: supported growth responsible for social stability, financial risk management, transition of an economy focused on

1. Laurent Leloup, *Blockchain, la révolution de la confiance,* Eyrolles, Feb. 17, 2017.

2. See Part I, Travail des données, nudge et réductionnisme.

exportation to an economy based on domestic consumption. The crisis of 2008 only reinforced the need to introduce a system of financial evaluation to the national level. To do this, China took much inspiration from the West, especially the United States. They are taking up the idea of a FICO score and learning from the mistakes of sub-primes. The relation and comparison to the American system which inspire it are often clearly indicated in the references in Chinese on the subject.

What about control of society? I think the SCS initiative should rather be compared to nudge politics. There are probably other ways of surveilling the population. It took the Snowden scandal for us to grasp the full magnitude of American surveillance operations. And we think that the PRC would unveil the same type of operation in broad daylight? Why communicate about it publicly if it's not essentially a deterrent? Thanks to the SCS, the individual is supposed to be aware of their behavior and improve it of their own will, or under the pressure of their peers. You might say, better the devil we know than the devil we don't.
But what do we really know about this devil? At this point, I must emphasize that my interest in understanding the Chinese system does not mean I endorse its functionality. What concerns me is how westerners, and especially journalists on a quest for buzz, dare to pretend they understand this beta-stage system, which is, incidentally, still fuzzy for most of the parties involved.

There is, first and foremost, the language barrier. I admit I fell for it myself at first, believing the system was *social* because it analyzed data labeled "social", alternative data on behavior and human values instead of financial data. If that were the case, the system would analyze the propensity to reimburse a credit according to one's daily behavior and their history on social networks. Especially as there have been examples, in China, of peer lending circles (P2P) which judged solvency according to the time and typing speed, the number of toothbrushes in a household, or even a person's beauty as a gauge of a bright future. And it's precisely because these social data don't suffice that credit defaults are widespread (not counting fraudulent or mafia-like platforms) and that a national system like the SCS becomes necessary.

In Mandarin, **"social" (社会)** means that the system affects the whole of society. It encompasses citizens, businesses, and institutions. The adjective qualifies the system rather than the type of data that feed the algorithm. The term **"system" (体系)** would actually be more correct in the plural, but Chinese linguistics don't make a distinction of number. Last year we counted about fifty pilot projects, born of private or public initiatives. There have surely been more since then, and we don't know yet which examples of good practice will be used to establish the final version for 2020. Some measures are already in place, separately, with ideally a sharing of cross-sectional data between entities. We are still far from the representation of an infrastructure of panoptic surveillance, a blend of Orwellian dystopia and the Eye of Sauron from *The Lord of the Rings*. A final point on the translation, the polysemous term **"credit" (信用)** reminds us, in our own language as well, of the correlation between our reputation as human beings and the way we manage our money – especially borrowed money. Knowing how to manage our finances has, for centuries, been associated with the idea of good morality.

In summary, the Chinese Social Credit System means

that for a good human in society – a largely capitalist society where it is good to consume – they must buy, even if it means borrowing, to contribute to the country's economic growth. It's an invitation to make the Chinese dream come true, a dream that looks a lot like the one we know in the United States, whose economic, cultural, and political radiance is now irrefutable.

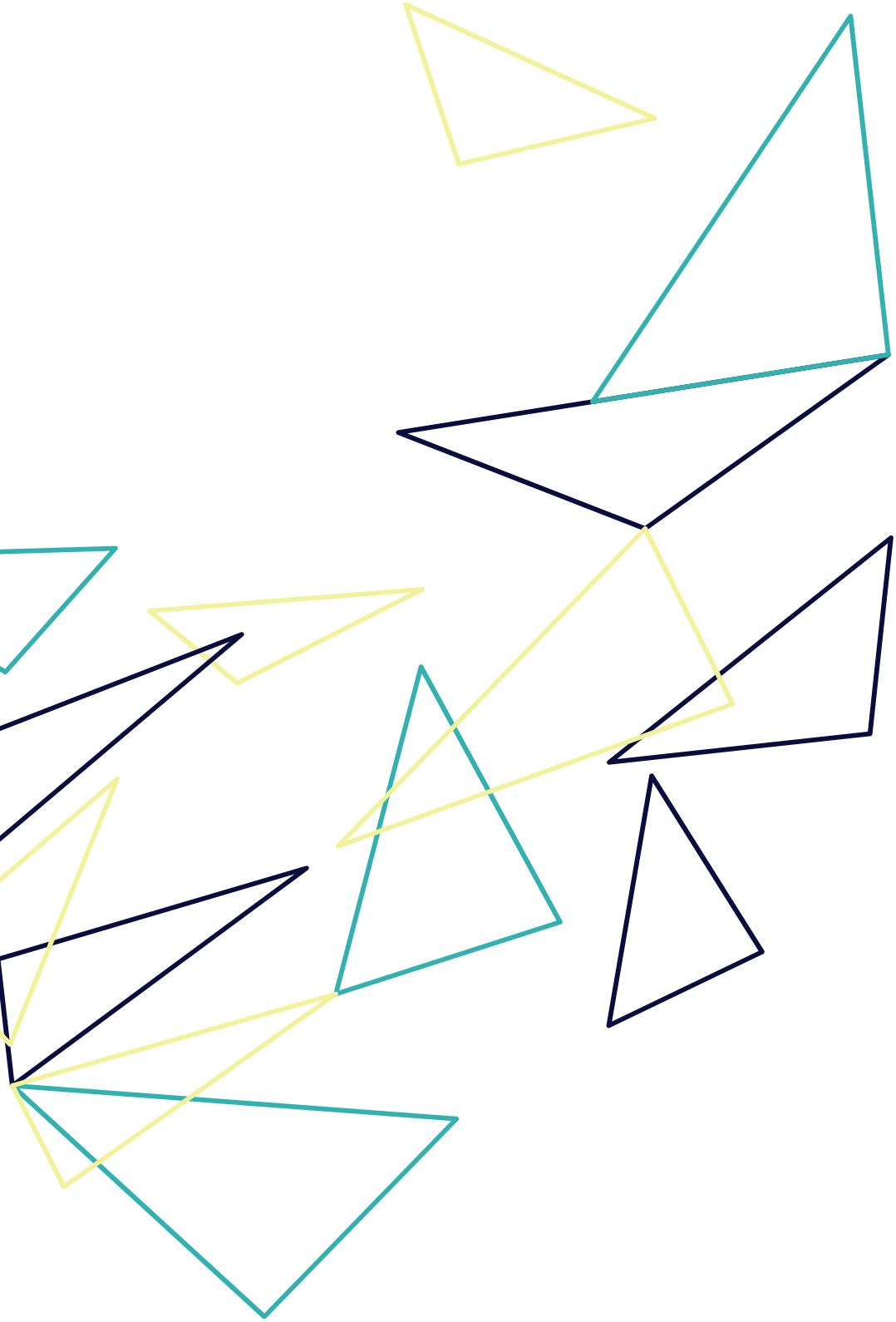
Admittedly, the SCS sins with an obvious techno-solutionism. The people, who are not fundamentally opposed to the system, seem to believe that the so-called neutral and objective technology will solve every socio-economic problem, from violence to corruption, from health scandals to rudeness on the highway. However, despite rapid technological advances, we are currently far from an intelligent, autonomous system devoid of all friction. The promises of artificial intelligence are not yet entirely functional. We talk about facial recognition to catch undisciplined pedestrians in the act. The reality sometimes shows us pedestrians within their rights who get whistled at by a police officer while a big hatchback does its best to run a red light, endangering the safety of everybody. The big screen at the end of the passage shows the same week-old image on a loop. There are two people, unrecognizable because they're bundled in their winter coats, all with poor image quality because of a thick fog. Human control has not yet fully given its place over to technology. A car does more damage than a pedestrian; what does our credit rating matter? I admit, the situation is evolving rapidly. In many ways, China is ahead of us; they could be a crystal ball to predict the future of those who know how to look into it: recognition of the harm of the internet and video games on our health (recognized by the World Health Organization ten years later, in 2018), prevention of hate speech, anonymity and misinformation on the internet for socio-political stability, society dematerialized with the digitalization of banking, legal, insurance, medical, or governmental services, etc., and soon perhaps, the digital normalization of the human being through massive data collection.

Are we able to judge what's happening in China, to influence its progress? No. The real question would be, how do we leverage this Chinese experiment in order to build our future? For us this Asian twin is a reflecting mirror, perhaps a magnifying mirror. What we criticize in Chinese society holds much more a projection of our fears than a current understanding of what's going on. The materialization of an omnipresent and omniscient system like what Orwell or Huxley imagined can trigger a deep reflection on the digital society in which we want to live.

We benefit from democracy, from a certain power to boycott and bring pressure, from the freedom to express our concern for the protection of privacy and the preservation of our integrity. Chinese citizens aren't complaining yet and not because they're being censored, but because the system is not yet in place. As for us, we should not wait until a similar system interferes in our societies to react. We can already lay the foundation with regulatory guard rails and a deep reflection on the definition of the human and their value in digital society. Are we more than the traces of our virtual activity[3], our financial and biomedical data? Can we tolerate that humans are complex, fallible, sometimes outside the norm? We already have trouble accepting the finiteness and the vulnerability of the earth and its resources….

3. See Part I, Individu, données, machine.

The SCS project in China is still in its test phase, and what good is an experimental phase if not to explore potential skills and tools, to wonder not only about what's possible with current resources, but also and especially about what is desirable? The energy spent on judging and pointing fingers should be catalyzed to mobilize us for the society we want. The debate should surpass the journalistic buzz that blindly praises new technologies and the criticism of everything related to "Yellow Peril". We cannot stop the march of progress, but we can hope to steer it. It will take a group effort to accompany citizens, empower businesses, and engage state actors to make a stand. As the cradle of critical thinking and scientific ideology in the 19th century, endowed with a strong sense of the individual, Europe must continue to play a big role in the world!



TOWARDS A CONNECTED DEMOCRACY

Prof. Helen Margetts, Professor of Sociology and the Internet at the Oxford Internet Institute.

"Social media are doomsday machines […] As a result, our social, civic, and political ligands are dissolving." Such were Jason Potin's defeatist words in Wired when we were reaching the end of the year 2018 and, with it, annual report season. His ideals of freedom of expression and technological progress, as for many other, suffered from the latest scandals. Even Tim Cook, Apple's CEO, reportedly admitted at a private conference in Brussels that platforms and algorithms is only giving free rein to the humanity's worst, instead of the best.

After each scandal, after each buzz, there are initial heated reactions. There was a wave of indignation after the Cambridge Analytica outrage, to name just one. Only few left Facebook after the debacle, though. How could it be explained? Prof. Helen Margetts, a political scientist specializing in governance in the digital age, offers us her thoughtful vision of the situation. But do not fear, all hope is not lost, far from it!
"I have not failed. I've just found 10,000 ways that won't work." (Thomas Edison)

The jury is still out, it is too early to decide. Generally, this is what I think when people tend to get too pessimistic about trends, says Prof. Margetts. There are things that we need to do instead, in terms of guiding the social media in the right direction, creating new regulations, educating people, and encouraging positive changes. There is so much more than focusing on what is not working well (it goes without saying that we should not deny it either).

There's so much negativity about the internet, it's true, which has the negative effect of stopping initiatives. People are more exposed to politics and engaging with it then before, and that is exciting. The point is how we are going to creatively and intelligently tap into that potential and create institutions that allow for low-cost political participation. I don't think a lot about the current balance of good and bad, admit Magretts. I think it's more interesting to think about how we can shape the trends that are forming, and about the fact that there is still room to shape them.

It is important to know that nowadays, if you talk to any tech scientist, they'll all say that a completely free internet is not possible. The very principle of freedom is to stops where the freedom of others begins, and for this to be guaranteed, there is always some king of regulation and control that must be going on.

We were convinced, at first, that internet would provide a voice for the disadvantaged and empower them. We believed that even in the most authoritarian countries, it would ensure freedom of expression. Democracy would have won by its own inner power. Internet was coined "the technology of freedom" up to the end of the 2010s. Nowadays, we sense mixed feeling and dashed hopes. Internet has been linked to control, surveillance, misinformation, and it is becoming a new war-zone.

Internet and social media are not inherently good or bad, pro- or anti-democratic. It is not any of those things. Digital platforms are intertwined with political life. We should think of them as merging or converging. One important thing that we should realize is that democracy is a small part of these platforms. These platforms are a big part of politics, but politics is not a big part of it.

Think about it for a second. You and I, to how many interfaces we are connected right now. Your phone next to you, Skype, emails getting in every two minutes, probably have Amazon open somewhere. Ten? Twenty interfaces? Probably all updated almost instantly. It ranges from professional to entertainment. In all of this, very little is political. With this in mind, the framework is laid for a more detailed look at the political life taking place in the digital sphere.

Tiny politics

As we go about our daily lives, for anybody who engages with social media platforms, this will modify political behaviour, it goes for politicians as well as for people. Donald Trump tweeting about foreign policy decisions, for example, shows how social media has become a part of politics. This has ramifications for how it plays out, for how many people talk about it. We are not all presidents, but even for average people, in most countries, if they use social media platforms, they are going to be invited to participate in politics. This might mean changes for individual citizen. It might mean that they can participate more. Some will

do so passively by taking an interest in current events, others will do so more actively, even if it is expressed in modest political actions. This is called the possibility of making "tiny politics".

Political participation is cumbersome. Historically, going to a march, joining a political party, all of this has high costs. It requires personal investment in the hope to bring about change. Internet undoes this. It opens the possibility of political participation with less cost, time and effort. Even giving a financial contribution, the size of the financial contribution. Before, the sum had to be greater than the transaction cost to send the it over, to be worth putting it in an envelope. Nowadays you can send money through text. There are a lot of little things you can do, you can follow a politician, sing online petitions, share them, express your opinions, donate, etc.

There are lots of ways you can do a little bit of politics. It has the potential to spur a snowball effect. We have seen this, the petition for blocking Trump from making a visit to Britain, it has scaled up. A number of large demonstrations, even revolutions like the Arab Spring, are the result of this scaling up.

The impact made possible by digital technology is exciting. Everyone can make a difference. Obviously, if you have heard about how in the US children campaigning for gun control has been made possible and that it has an impact, that's exciting. That are so many examples of mobilizations that are challenging injustice. Another example is the Romanian people, and how they have recently organised to denounce corruption and held their government accountable. One interest thing, Romanians outside Romania have played a big role in the mobilization, which is a completely new element. In another time, they wouldn't even by a player in the political system. Once again, the Arab spring, the Brazilian protests, it would be difficult to point out a country where internet hasn't played a role recently in allowing for actors to organize.

We see the diverse nature of these modest political acts amplified by technology: sharing a photograph of a refugee child, or a piece of misinformation, for example about the European Union organizing its own army, they can also scale up to the point of altering the political scenario. Two examples illustrate these unforeseen developments. One is our current unlikely leader of the labour party, Jeremy Corbin, who is a political with a very particular background that has received a big wave of support [in 2015], and it comes from social media. The Barack Obama campaign is another good example, in which it received a lot of money but from small contributions, and it was the first mainstream campaign to do this.

By doing something with little cost, some people may criticize this idea, saying that is insignificant, calling it "slackativism", but I think these are still important, and they mean two key differences for politics.

1/ First, this practice brings visibility. When you do this type of politics, the actor, the cause supported and its audience are made relatively public. By talking to people in a bar, you may talk to a couple of people, reach a couple of people. By liking, tweeting or sharing, you'll reach more people than in a bar, probably.

2/ Then, this practice brings amplitude. By sharing, or linking, you are sending a little signal to other people, insignificant at first glance, but if enough people do it, it can scale up to something big. What seems to be just a bottle in the sea of online activities can grow into a wave of reactions.

However, there is no distinction between good and bad content as both can grow in scope. Hate speech, misogyny or racism can also scale up. Mostly it doesn't, but it can, and the impact of these scaling up is affecting political equality.

The real problem: uncertainty

A third key difference should be added, and not the least. This practice brings uncertainty. It is a worrying factor of tiny politics. As such, tiny political acts can have positive or negative dimensions. Most of these tiny political acts don't go anywhere. Most pieces of fake news don't scale up. We see big mobilizations or other actions that succeed and it makes us think that it's easy, but it's not. When we do manage that, why some campaigns are successful and others are not, why something creates a wave of support and other things do not, we don't actually know. Tiny political acts have a major influence in the sense that they inject randomness to politics. This adds uncertainty and instability to politics at the moment. Uncertainty, almost by definition, is something negative. We have this idea that political institutions have a stabilizing effect, that they exist to iron out uncertainty. What is worrying is that the uncertainty that exists regarding institutions and their regulation of digital technology, or the lack of institutions, can turn political systems chaotic. They are now all being challenged by large scale popular mobilization, and by the fact that corporations behind social platforms are very important in political life.

"The best way to predict the future is to invent it. " (Alan Kay)

It is difficult to predict future trends. What we can do, what we should do, is to improve infrastructure through rules, design and education.

Rules, Design and Education

Interface and platform designers are the first to be able to make a difference. Design has an impact on users, their behaviours and choices. Thaler was awarded the 2017 Nobel Prize for demonstrating this with his analysis of nudging. Choice architecture, while offering a certain freedom, can lead to a more favourable result than another. This does not mean that the platform design allows users' opinions to be manipulated.

How platforms are designed impact our politics, our sensibility to issues. In the elections for instance, internet did not change people's voting per se. Rather, it changed the decision of some people to vote or not, as a result of facilitated access to political life. This factor partly explains the election results. Facebook had tested its algorithm to find out what settings people were publishing on their votes. When the platform lets you know what people are doing, what information the platform gives you, knowing how many people are touched by an information published by the platform, that's what Facebook, Twitter or Instagram do, and snapchat, on the other hand, doesn't do, this is an important element in creating impacting. People tend to see differently information that has touched many people, and information who hasn't. Choices conform or are reinforced by social pressure. This technique called in psychology the "normative influence" has been used for a long time in marketing. But I think these dynamics are too recent, and they are still taking shape and form, and there is still a lot of possibility for us to stop and evaluate now. We still have time to institutionalize it, to establish design rules.

All responsibility does not lie solely with the designers. The user who can make a difference. Citizens could learn to better understand new technologies

and their use. The benefits would extend beyond the framework of political life. In Sweden, an initiative invites the whole population to be trained by MOOCs in the basics of artificial intelligence. Citizens would thus be equipped to better understand the society around them, to better distinguish the issues at stake during referenda, in short to better participate in democracy, not to mention the prospects for employability.
Where education can make a difference is to distinguish misinformation, to seek the origin of information, and to develop healthy scepticism. People can also be trained in their rights and the tools to maintain control over technology. It is essential to understand how these systems can work. In the case of Russia, we talk a lot about computational propaganda, but we don't know a lot about it. To what extent do people really understand? We have a lot to do, and we have to get better at it, but I don't think is the end of democracy in the digital age.

The road to numerical hell is sometimes paved with good intentions

The growing dependence on social networks over a traditional information medium is often criticized. It is believed that because of personalized flow, individual's opinions are reinforced. social networks that are constantly testing their algorithms to find the right balance between centres of interest and new perspectives. Social networks don't want to be a political actor. They are not trying to change democracy.

They want to make a lot of money on making people connect with each other and spend a nice time. Their business model has been damaged for the role they have come to play in recent political processes. For us, that is a win, because this damage to the corporation will make it more willing to cooperate to make democracy benign. So, they are engaging afterwards, hiring people to clean up and control the content. It is worth mentioning that these dangers can be even bigger in the small countries, Facebook has taken up to cleaning its platforms, but it does it mostly in English, and concentrates on users located in central countries. Which means that damaging dynamics in other countries go under the radar, the case of Burma made this clear.

Facebook has learned the hard way about the importance of not locking the user in a bubble. The platform had this problem with trending information, because it was made by people, and so there was no hierarchy of information, and this was a problem. Then it was organized by robots, and this also became a problem, and people would say Facebook is manipulating us, so now Facebook took the feature down altogether. But this actually makes us less exposed to all kinds of different information, and we see less. We do have to get better. Social media should be encouraged to develop appropriate trending functionalities.

We have made social media into a demon. Nothing gets solve by doing so. Closing down a platform wouldn't help; people might be using something that is worst. You can't make people go where they don't want to either. One direction ahead is to demand improvements from private actors. It is debatable to what extend we can ask these companies to be arbiter of what is good and bad media. They are not media in the traditional sense. Donald Trump argued that social media is biased against him. Corporations do have responsibility for the content on their platforms. It shouldn't be a fake news farm or a Russian bot, and it shouldn't be hate speech, but how far do we want

to go in the road of them saying what is and is not news. Of course, this kind of regulation can be bad if it's China. But, on the other hand, Germany, for historical reasons, we would understand greater regulation on hate speech, and internet regulation would be used in positive ways.

A new art of governing: equality, equity and quality

In fact, this article is a call for creativity. Think about all the new ways of how digital platform allows us to interact in a low-cost way. It is exciting that more people participate in political life, and that more people know what is going on. Someone with nothing more than a mobile phone can take place in political life. Refugees fleeing Syria will prefer keeping their mobile phone to food, to a blanket, to anything, because it gives them more autonomy, it allows them to contact family, connect with people, register their journey and situations they encounter.

On a different note, I am thinking of British politics. In Great Britain, we are one of these countries where you have electoral systems based in small constituencies, and we carry on with these electoral systems because of this idea of constituencies, because of the idea that these constituencies allow for closer relations between the representatives and the people they are representing, and this is actually rarely the case. It is rare to find people that are actually involved and in contact with their representatives. Internet might allow us to rethink this electoral system that doesn't work in the way expected and actually create new ways for the representative democracy to work.

It might be that you encourage more people to participate in political life, but I don't think that we should go to direct democracy. We had this ideal at the beginning of the internet. Yet being theoretically feasible though digital technology does not necessarily means being desirable. We have to re-invent representative democracy. That's the point. We have the possibility for creating other forms of the State to engage with citizens. That's more promising. The State would try to interact with citizens more frequently and differently. Proximity and dialogue would build trust. People don't trust Facebook, and they don't trust State either. Early trends are rather positive. We must take advantage and continue to improve communication and political commitment. Whether for politics, health, sustainable development, education or any kind of public service, the Internet opens up a field of possibilities to be exploited.

This would also demand more sensibility from the State. It must be able to identify services that could benefit from a digital transition, but also the people who are familiar with technologies. It means expanding opportunities to those who require fast and remote solutions, while allowing other to be treated differently if they want to continue with the current method, out of necessity or comfort. It is difficult, because the State is based on this idea of seeing all citizen as equal.

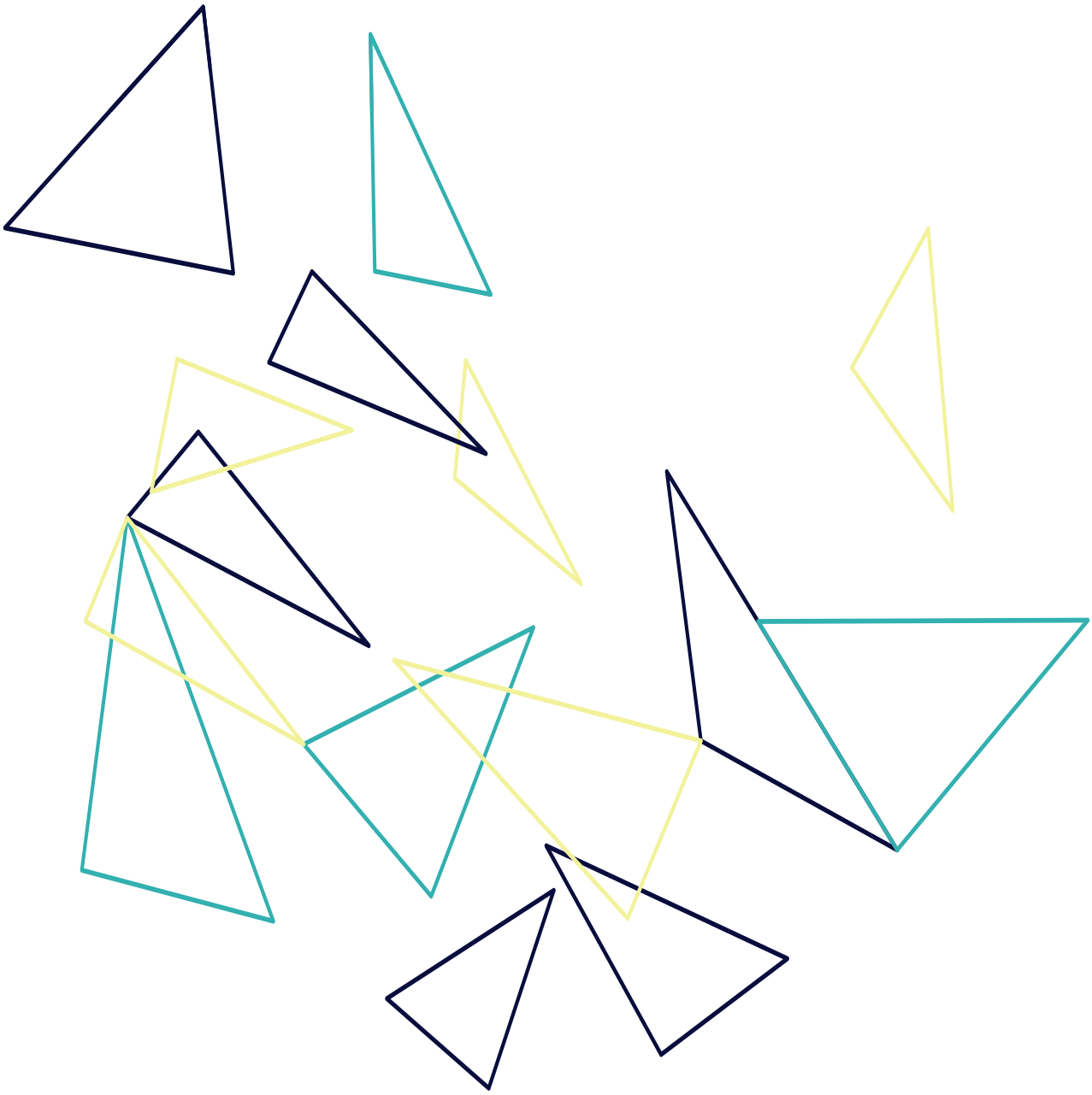
How should the State engage with its citizens? Which means should it use? How to manage public-private partnerships? We need to think about creating a different and more frequent, and citizens would be able to express their opinion through these platforms. The State could collect more data on how citizens think and feel about policies, either in their design or application. It might enable to identify sooner failing ser-

vices, like schools or hospitals, because people would be talking about it. I think that States have to try to get better at collecting data, even if the idea, linked to the fear of surveillance, is controversial.

The State could try to finding out what people think and want. The Platform State would be more reactive and citizen-focused. Similar to their commercial counterpart, they would be service-oriented, except that the clients here are the citizens. We would have much more to gain from such changes. Maybe, it could be more like soliciting people, considering platforms present fantastic opportunity to understand and evaluate how people experience policy change. The general malaise of the latest scandals should not slow us down. On the contrary, there's a lot of catching up to do. We should learn from the mistakes of both the private and public sectors to constantly to improve.

The recent dynamics around digital technology are both exciting and worrying. It is still too early to know on which side the scales would tilt. Let's not forget that technologies are man-made. It is up to us to shape them for the future we want to build. This future, with a freedom of expression that has found a new place on the Internet, must be filled with tolerance to allow the society to function despite our differences.

Freedom of expression to everyone without supervision is not the same as democracy. To work well, this tolerance also requires a legislative framework. The current platforms are not machines of truths. This may not even be what users want, knowing the importance of entertainment. People are more interested to participate in political life that we though. The importance of tiny politics is no way less, it has shown positive impact in making things happen!



ETHICS & TECH 2019 Report

Part III
RUPTURES & CONFLICTS

The apparent or real complexity of technological subjects is also in itself a source of mistrust. The classic "black box effect" can lead to uncertainty, suspicion and mistrust when the problems raised by these complex systems and innovations reach audiences through uncontrolled media coverage, the objective of which is often to produce a side-real effect conducive to "buzzing" and "clicking" by soliciting the imagination.

In recent months, we have identified two themes – among many others – that have largely fuelled the collective imagination, leading to social metamorphoses. These two themes, which illustrate the acceleration of changes in the imagination produced by digital technology, have recently been met with forms of disillusionment and mistrust: **they are blockchains and the militarisation of Artificial Intelligence.**

In both cases, this acceleration, specific to the digital transition, produces new problems linked to a fantasy of hyper-rationalisation and hyper-efficiency that advocates the progressive erasure of human beings with our uncertainties and weaknesses.

Bitcoin, as an application of blockchain technology, caught the imagination through the hypothesis of a crypto-monetary asset supposedly free from the control of political power and trust and finally achieving the anarcho-libertarian ideal. The other example is the creeping, increasing militarisation of AI in conflict zones, which has fuelled the hypothesis of a digital apocalypse, a loss of arms control and reactivation of the original myth of dog-eat-dog.
In both cases, collective and reciprocal trust seems to be seriously compromised.

In the following pages, we would like to go beyond a superficial and anxiety-causing approach to these two phenomena, to show that on the one hand, blockchains as a protocol can, contrary to the philosophical tendencies that gave rise to "crypto-currencies", initiate a form of neo-mutualism, using the interesting case of future governance of business ecosystems collectively "orchestrated" by solutions resulting from this technology.

Secondly, the field of war and conflict must be approached with pragmatism and yet hope. Like nuclear technology in its time, it would be illusory to think that the pursuit of political interests itself refuses to use new technology. Nevertheless, and putting cynicism aside, it is clear that we are on the cusp of a period that will require a deep desire to limit and control this new warrior fantasy, to the point perhaps of gradually and collectively formulating the idea of a Peace that is as digital as the wars to come.

With the participation of:

Damien de **CHILLAZ**

Edouard **MORIO de L'ISLE**

Dominique **LAMBERT**

Marylaure **BLOCH**

Pierre **GUEYDIER**

Romina **REBOIS**

Mariarosaria **TADDEO**

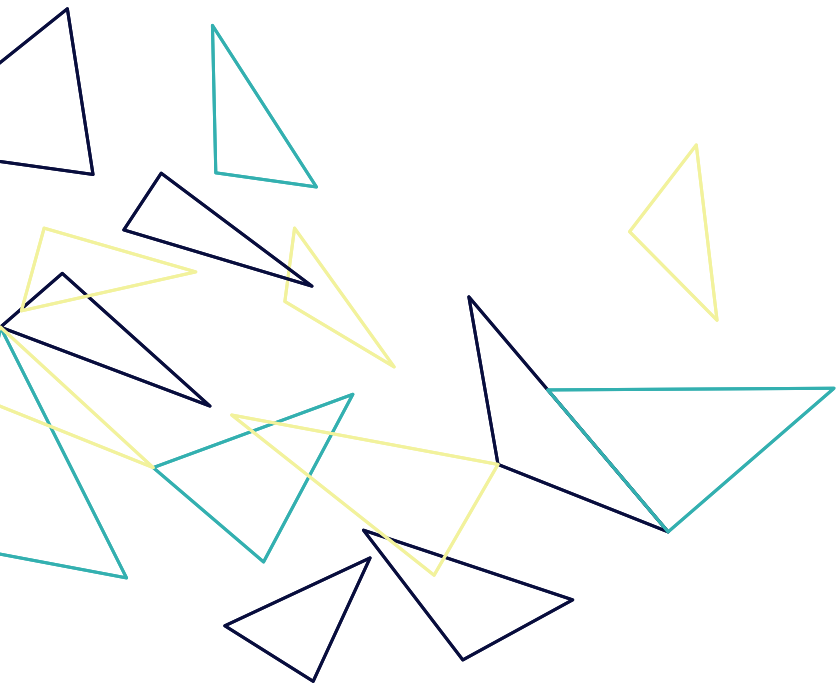
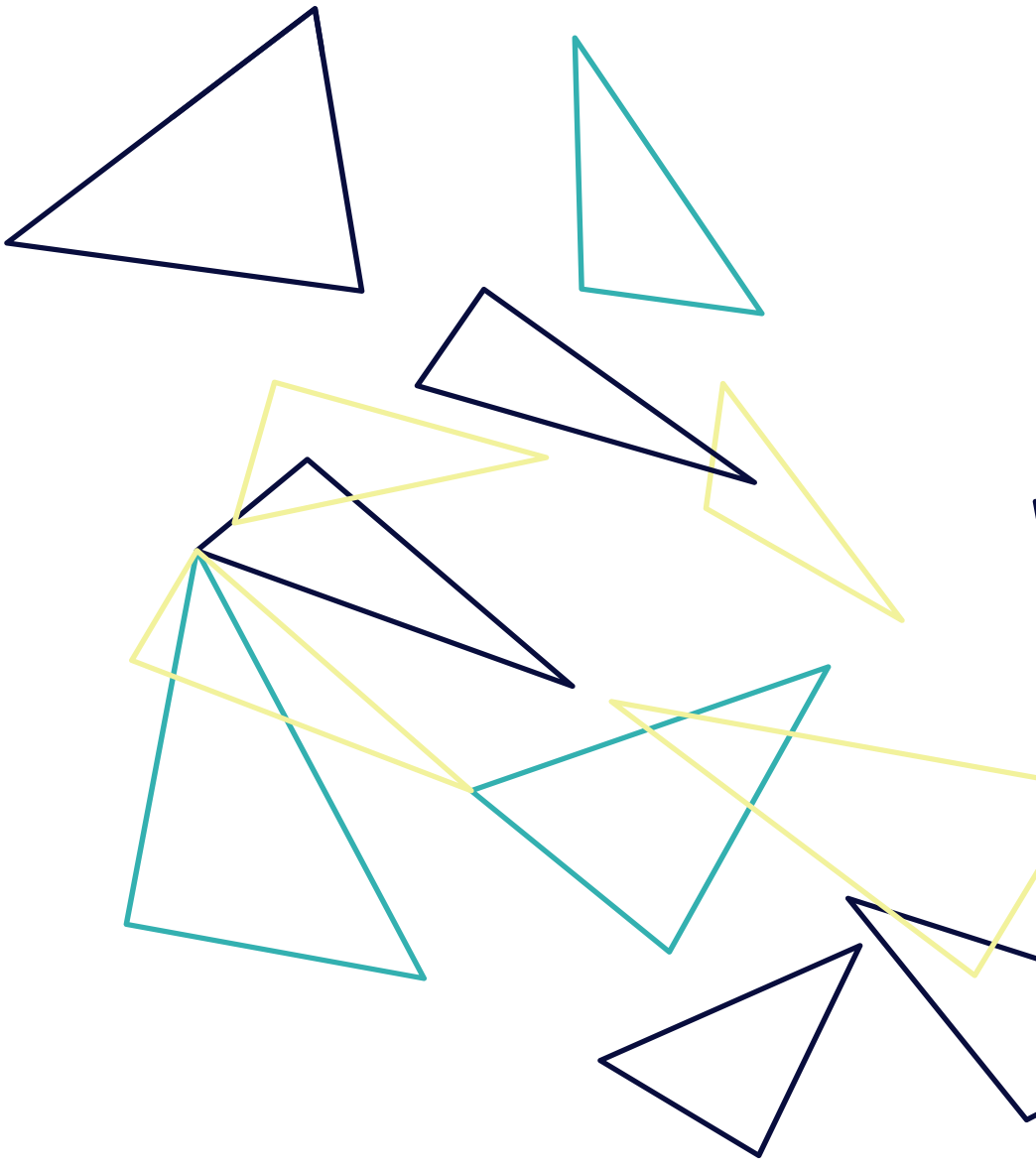


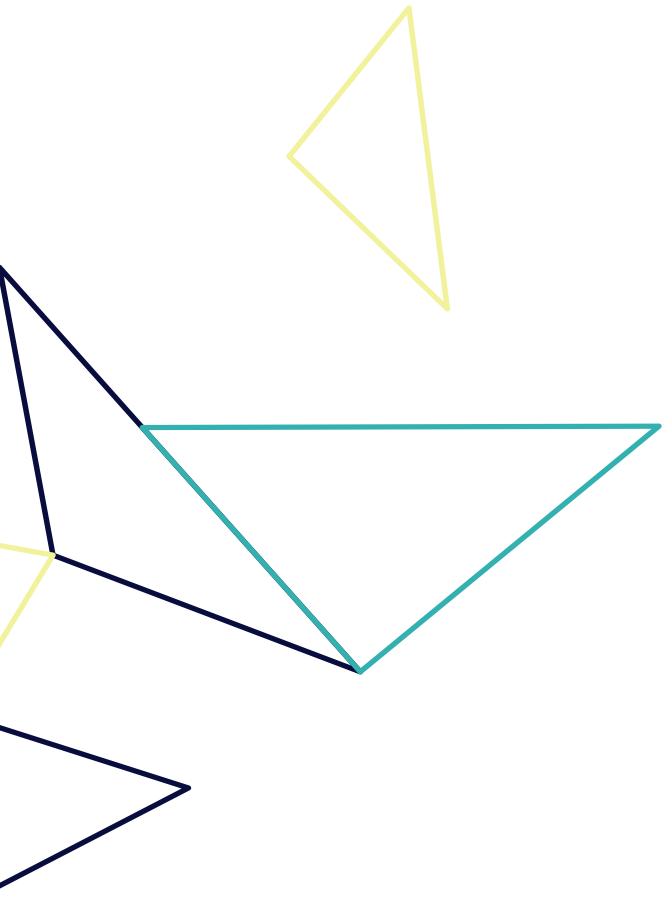
TABLE OF CONTENTS



Blockchains: damaging or conducive to trust?

Uncertainties, collapses, trends, controversies, manipulations... 2018 represented all of this and more for "crypto-currencies" and their underlying technology: the famous "blockchains". If nothing else, these developments will have at least had the merit of bringing this curious innovation into an exciting and destabilising discussion space. Indeed, they bring into brutal question a cardinal value of human relations: that of trust as the cement of the social contract.

While institutional actors (States, central banks, international institutions) now seem to be jockeying for various positions with regards to this new technology, a question arises for the business world: can blockchains renew a mutualist approach to B2B relations? In other words, can blockchains foster trust and the search for the common good among the various economic actors?



Overview of blockchains in 2018.

Early 2018 difficult for crypto-currencies

The first few weeks of 2018 saw the price of Bitcoin hit the headlines with a sudden collapse following the spectacular surge in the last few weeks of 2017. Between the end of December 2017 and March 2018, Bitcoin, which also brought down all the "crypto currencies", depreciated by about 70% in three months.[1] This was the fourth historical decline in these assets since they first appeared in 2009.

The extreme volatility of these assets is naturally based on fluctuations in stakeholders' confidence levels. The bursting of this speculative bubble is not unlike the one that struck another famous protocol in the late 1990s: TCP/IP, better known as the internet. But the media focus on "crypto currencies" and their fluctuations should not prevent astute observers from taking an interest in the related phenomenon of "Initial Coin Offering" (ICO) in recent months.

This method of raising funds by issuing digital assets ("tokens") exchangeable for "crypto-currency" and based on blockchain technology was launched in 2013. Although they are based on the classic "Initial Public Offering" (IPO) method, which proposes the acquisition of shares in the capital of a company, ICO tokens do not represent capital shares and operate within a legal framework that is, at best, vague.

Despite the total absence of any guarantee for "token" buyers, ICOs are presented by their promoters as a breakthrough in the investment world that is supposed to break down barriers between professional investors and individual buyers. Just as blogs were able, in their time, to propose the same breakthrough into the world of professional journalism. Thus, 2018, which began with the destruction of the value of "crypto-currencies" that had "flown too close to the sun", continues with a worldwide craze for ICOs as a means of financing a galaxy of start-ups.

Legal uncertainty and volatility are not the only weak points of the "crypto-currency" universe. The last few months have also been marked by scams and malicious acts, the frequency and extent of which have increased over the months. These include simple phishing scams to attempts to illegally recover assets. When these offences are suspected on trading platforms, these platforms are generally obliged to block all transactions as a security measure. However, this type of precautionary measure can very quickly trigger panic behaviour and brutally affect the price of assets and the reputation of a particular actor. More seriously, some criminals are reviving old methods of robbery, either using violence against people to force them to transfer their assets, or, quite simply, to steal property through the physical theft of servers used to undermine "crypto currencies".

2018, central banks and regulators address the issue

1. "L'année 2018, marque l'éclatement de la bulle des crypto-monnaies", *Les Echos,* 27/12/2018, https://www.lesechos.fr/finance-marches/marches-financiers/0600402112585-lannee-2018-marque-leclatement-de-la-bulle-des-cryptomonnaies-2232669.php (in French).

Not surprisingly, in the face of these destabilising effects and promises of disruption in the investment world, the historical players in the field – central banks and public regulators – have brought this phenomenon to the attention of their legal and prospective experts. 2018 and, even more so 2019, will be a turning point in the way in which the financial applications of blockchains are managed.

The first half of 2018 was therefore one of an abundance of reports and position papers of all kinds. Crypto-currencies, smart contracts and ICOs received the attention of the world's main players and regulators, from central banks to the IMF, from the World Bank to the Bank for International Settlements (BIS), the real central bank of central banks. By means of evidence of the concern of major financial institutions, various studies have led their leaders to take a public, and sometimes brutal, position. Thus, in February 2018, Jim Yong Kim, the head of the World Bank, compared "crypto-currencies" to Ponzi schemes, while acknowledging that "we still do not know clearly how this will work" and that beyond the "crypto-currencies", blockchain technology could have positive applications to "follow money more effectively", thus reducing corruption.[2]

In April 2018, it was the turn of the International Monetary Fund (IMF) to express its position. While stressing the risks, the IMF was more nuanced and suggested that, in the long run, beyond crypto-currencies per se, "distributed ledger" technology will be more effective and could have many logistical advantages for international financial exchanges and the security of all kinds of sensitive data.

The G20 Financial Stability Board published its recommendations in July 2018, concluding that, despite the risks, there was no systemic risk regarding crypto-currencies and that States and regulators should coordinate their analyses and positions on the issue.

An abundance of public statements by central banks was also noted in the first months of 2018. Indeed, since the overall objective of blockchain and crypto-currencies is to challenge or even destroy the centralising function of monetary policies, their opinion is – not surprisingly – generally negative.

However, despite this reticence and the lack of consensus on the analysis of the phenomenon, the positions of the major global financial actors coexist alongside the opportunity for sovereign state actors to take privileged positions in the event that the economic applications of blockchains prove to fulfil their promises in the medium term. This is the case in France, for example, where the very reserved opinion of the Banque de France contradicts that of the Autorité des Marchés Financiers (AMF), which suggests that competition between European countries for leadership on these issues is likely to begin, particularly in the specific context of Brexit. The precipitation of

2. "Le patron de la Banque mondiale dit que beaucoup de crypto-monnaies sont des systèmes de Ponzi", *Business Insider,* February 2018, https://www.businessinsider.fr/jim-yong-kim-banque-mondiale-crypto-monnaies-pyramide-ponzi (in French).

small states, which have traditionally based their power on the financial industry, is also an indication of an evolution that is no longer limited to the anecdotal aspects of Bitcoin. Thus, on 4 July 2018 the Maltese Parliament officially established the first regulatory framework for blockchains, crypto-currencies and DLT (Distributed Ledger Technology), making Malta the first country in the world to provide an official set of regulations for blockchain, crypto-currency and DLT space operators.

Finally, from a more geopolitical and conflictual point of view, several States subject to international financial sanctions are studying the possibility of creating a centralised and sovereign "crypto-currency". One of the first projects was carried out in December 2017 by Venezuela, which was facing unprecedented inflation. Petro, a crypto-currency asset backed by Venezuelan oil, was officially launched in February 2018 but faces numerous operational obstacles, which did not prevent the Venezuelan authorities from announcing a new currency backed by this crypto-currency in August 2018. In the same vein, Iran, also facing very high inflation and drastic international sanctions, has announced the launch of an experimental "energocoin" backed by its energy reserves. Russia, for its part, is announcing such measures for 2019.

To conclude this brief overview, it should be remembered that 2018 is a paradoxical turning point for blockchains and crypto-currencies. First of all, the phenomenon received very strong media coverage during the hyper-speculation at the end of 2017, which contributed to popularising the theme. The collapse of early 2018 could have been the end of the subject but, at the same time, the main financial authorities, while stressing the colossal risks, launched and made public vast studies that are proving highly indecisive and controversial on the academic assessment of the phenomenon. Finally, 2018 will have been the year in which the inter-state struggle began to adopt future positions on this new source of power and sovereignty. In other words, while "crypto-currencies" have monopolised attention to the point of often being confused with the protocol underlying them, we must take care not to lose sight of the wood for the trees but, on the contrary, place blockchains, as a technical protocol, at the centre of attention in order to evaluate their potential for uses in conformity with the Common Good.

Blockchains and platformisation of the economy

Towards a generalised platform

The most promising scope for blockchains lie in their connection with another fundamental phenomenon in the evolution of economic exchanges, that of multifaceted markets, popularised under the term "platform". "Platformisation" has long been clearly theorised by economics. This mode of optimising trade, which dates back to the invention of the "marketplace", is ultimately linked to an information economy process that has been profoundly accelerated by the digital revolution. The ability to bring together large masses of economic agents in the same place (physical in the case of the marketplace, now digital) makes it possible to increase the externalities of these networks of agents through reduced transaction costs. Transaction costs (time spent searching for the product, price comparison, etc.) are essentially internalised by the platform. This crucial significance of information in price structuring has therefore enabled the emergence within a few years of "giga-platforms" with strong bilateral market power: they decide the price charged on either side of the market.

Few sectors escape this heavy positioning and the concentration of economic power in the hands of these platforms has never been so important. Consequently, the trust that users place in these huge economic actors is crucial. All the more so because, as we will see, this phenomenon is only in its infancy and the future of the digital-age economy looks to be that of interconnected "giga-platforms".

Emergence of the concept of the digital ecosystem

Observation of business ecosystems has gradually identified the crucial importance of the digital platform as an infrastructure for innovation dynamics. To coordinate a digital ecosystem, it is therefore crucial for the "lead firm" to manage the tension between collaboration and competition through a platform strategy, as well as to implement its governance and architectural choices. The leader of an ecosystem will therefore generally be the organisation that has identified and implemented the most beneficial collaborative arrangements for ecosystem members, resulting in collective prosperity and survival. The image of a musical conductor is useful here to describe how, by controlling architecture and governance, collective value can be created by pooling and sharing resources.

It is worth noting here that these reflections, in attempting to go beyond traditional theories of competition, seem to give way to behaviours that are open to forms of mutualism in economic relations or at least to new forms of economic power that are more widely distributed and open to "multi-stakeholder" governance.

In the 1980s, proponents of multi-party governance structured their initiatives in response to the failure and ineffectiveness of self-regulatory private sector policies. These initiatives sought to revive the long-standing values of subsidiarity and the informed and active participation of stakeholders in a collective problem or issue that produces negative and positive externalities. The central concept of subsidiarity, coupled with that of trust, has always been promoted to avoid abuses of power, whether they stem from excessive state control or from the violence inherent in the principle of economic competition.

Governing a digital ecosystem: the blockchain hypothesis

This raises the essential question of the governance of these digital B2B ecosystems. The technological architecture must include principles that are conducive to the organisation of virtuous interactions between ecosystem members. Indeed, technological choices are never neutral but reflect the fundamental principles of governance through which the future interactions of ecosystem actors will be able to engage harmoniously.

The viability of these new organisations therefore lies in the ability of a "conductor" to generate trusting relationships between members of the ecosystem. The principle of trust is not new in business relationships, but it is crucial here, given the uncertainties created by these rapid changes in economic paradigms and the scale of their impacts. As a result, innovations in business models and technological architecture must also lead to a profound revision of the principles of governance.

Beyond these generous ideas, which may often be utopian in the face of the reality of human nature, the question is the following: are these new information systems capable, through their architecture (platform) and their trust protocol (blockchain), of supporting such virtuous large-scale governance projects?

Experiments are under way to develop a governance of interconnected platforms coordinated by one or more private blockchains capable of ensuring an optimal level of trust between stakeholders. For example, the regulatory issue of vigilance regarding the identity and nature of client activities in the world of banking and finance currently provides an excellent opportunity to test new modes of governance and balance of interests. Indeed, this regulation concerns all companies seeking to enter into a relationship or already in a relationship with banks. A project to "platformise" this process within a global network where each company could exchange its information in a consensual, secure and traceable way thus immediately faces a major challenge of governance. The aim is to find a model that promotes trust in the technical and functional infrastructure of the platform but also in the overall governance between the different members of the network. In this specific case, the aim is to create the conditions for the emergence of trust between these stakeholders. The objective is to potentially encourage any company to join this network. This network is therefore intended to be global and inclusive, which will make it possible to promote its adoption and the gradual creation of the famous "network effects".

At this stage, it appears that blockchain, and in particular Corda, as a corporate blockchain protocol, responds satisfactorily to the technical governance of this platform, particularly in terms of security, confidentiality and traceability by allowing the bilateral exchange of information within the network. However, it is not in itself a response to the corporate governance of this network i.e., the way decisions are made about the rules of engagement on the platform, such as the process of accepting or not accepting new entrants, which is done according to a number of criteria that must be defined by the decision-making body that literally "governs" the network. The challenge lies in the definition of this higher body, particularly in its legal form and its internal decision-making process. These choices, whether technical, functional or governance-related, must be focussed on assisting the members of the network in order to serve the Common Good.

It is clear that technology alone will never completely erase any company's responsibility to define its mission, values and operations and to assume the consequences, whether economic, social or even political and environmental.[1]

Conclusion - Towards a neomutualism with blockchains?

The main positive outcome of the emergence of blockchains and "crypto-currencies" was to produce a collective awareness of the essential importance of technical infrastructures in political processes at the end of the 2008 financial crisis. Nowadays, the debates surrounding governance issues and their links with technical architectures are being taken into account by numerous new audiences.

The violence with which the anarcho-libertarian movement has tried to promote the substitution of the virtue of trust of any social contract by a technical artifact has stimulated, and continues to stimulate, collective reflection. This form of techno-prophetic overkill has contributed to the emergence of the beginnings of reflection on the proper use of these techniques and their possible contribution to the Common Good. As we have pointed out, a protocol is never neutral. It not only serves an operational goal of efficiency, but also contributes, in the background, to the production of collective social utility. If it were needed, blockchains could underline the accuracy of Bernard Ziegler's definition of any technique such as "Pharmakon": both a poison and a cure but also a potential scapegoat for political negligence.

Through recent events, it is quite clear that this technology does not yet provide a satisfactory answer with regards to the Common Good. This is probably an initial state, the deficiencies of which will have to be corrected to prevent a sudden return of politics. Nevertheless, blockchains, particularly regarding their acceptance in the private sphere, are gradually emerging as an architecture capable of producing, on a large scale, the operational bases of shared trust in areas where it is necessary to coordinate multiple stakeholders within large digital ecosystems.

The good news in recent months is that these collective problems are being debated and questioned more and more, especially in the public sphere. Thus, the digital economy that is based on the exploitation of personal data is beginning to reach a kind of collective tolerance limit. In this sense, perhaps we can imagine that B2B experiments with blockchains in vast multi-stakeholder ecosystems will give new credibility to the concept of mutualism as an implementation of the principle of subsidiarity.

In any case, these experiments already show that the technical governance of blockchains, whatever the protocol used, cannot replace the corporate governance inherent in any professional organisation, and the responsibility that goes with it.

2. Details of this project appear at the end of this chapter in our interview with Damien de Chillaz, Vice President B2B Platforms & New Business Models at Capgemini.

To find out more:

Texte OPTIC, Blockchain, au défi de la confiance, http://www.optictechnology.org/images/files/Research/OPTIC2017-Blockchain-au-dfi-de-la-confiance.pdf

Maël Rolland and Assen Slim, "Économie politique du Bitcoin: l'institutionnalisation d'une monnaie sans institutions", Économie et institutions [On line], 26 | 2017, https://journals.openedition.org/ei/6023

De Filippi Primavera and Benjamin Loveluck (2016), "The invisible politics of Bitcoin: governance crisis of a decentralised infrastructure", Internet Policy Review, vol. 5, No. 3, September. https://policyreview.info/articles/analysis/invisible-politics-bitcoin-governance-crisis-decentralised-infrastructure

Towards an economy of mutuality with block-chains?

Interview with Damien de Chillaz

Vice President B2B Platforms & New Business Models at Capgemini

Can you describe the KYC Trust project to us?

This project involves the creation of a digital platform for companies, known as "B2B", linking banks and their corporate customers in order to allow the exchange of information necessary for the relationship- building process. This process is commonly referred to as "KYC" (Know Your Customer), hence the code name of this project, which also emphasises the central value of trust in these business relationships.

In a very practical way, this KYC process begins with the bank sending a questionnaire to its client, including around 300 questions, leading to the collection of data and documents providing information on the company's activity, its directors and shareholders as well as its legal organisation. This information is currently collected by e-mail, in a non-secure manner and partly based on centralised databases that have gathered some of this information. The poor reliability and quality of this information is a problem for banks, as is the cumbersome nature and redundancy of the process on the business side. In particular, these businesses complain that they do not properly control this exchange of information within their group, and that they have to respond several times to similar requests from their banks. There is a general consensus that the current information-gathering process is inefficient, costly and time-consuming, and is a task with very little added value for the bank or its clients.

Our KYC Trust project aims to give companies back control of their information, allowing them to exchange it simply and efficiently with their banks. To achieve this, we propose the creation of a global platform, based on a digital network secured by blockchain, and governed in such a way as to guarantee the integrity and neutrality of this network. The primary objective is to create the conditions for massive adoption of this global network, which will gradually reflect the degree of trust that each economic actor can have within its own business ecosystem. The bank-corporate relationship around KYC is therefore only a first use case, a first entry point towards the creation of this network of trust.

How does the technical architecture of this platform contribute to creating this trust?

This platform is designed in a completely innovative way as a secure digital network, linking each company to its banks, each actor having a "node" of the network allowing it to store its information, and to exchange it in a secure way according to the banks' regulatory requirements. We used the Corda blockchain technology developed by our partner R3, a consortium of more than 100 financial institutions, to build the technical architecture of this private network. Unlike

other blockchain protocols, Corda allows information to be exchanged "bilaterally" and it is not more widely distributed across the network. This "hybrid" blockchain protocol is therefore perfectly adapted to the needs of exchanges of sensitive information between legal entities, which are the elementary molecules of our network. Each legal entity, subsidiary of a company, or bank is uniquely identified by a 20-character code which is specific to it, the LEI (Legal Entity Identifier), which guarantees the uniqueness of this entity. In addition, we uniquely identify the people who are responsible for these information exchanges within companies (treasury teams) and banks (compliance teams). The platform therefore makes it possible to properly identify the persons and legal entities involved in the secure exchange of this sensitive information. Finally, we would like all fiduciary information circulating on this network to be e-signed by these previously identified persons in order to best guarantee their integrity.

It is clear that the scope of this network goes far beyond the limited framework of KYC banking and more broadly supports the exchange of information within each business ecosystem made up of banks, but also and above all of customers, suppliers and other stakeholders essential to the life of the company. We are therefore at the heart of this "economy of mutuality" in which trust between players is essential, and can now be objectified thanks to these new technologies if they are oriented towards the Common Good. Technology, essential though it may be, is not, however, enough to guarantee this Common Good and to limit conflicts of interest between regions and actors in an economic environment of global competition, where access to data and its use become the real source of value creation and competitive advantage. This network must therefore have inclusive, independent and autonomous governance.

What can you tell us about this innovative governance model?

Given the global nature of the banks and multinationals that will be members of this network, we have designed this platform to be global, and to accept entities of all sizes, sectors or regions in an inclusive manner. This is an essential condition for the successful adoption of this platform, and one of the fundamental values of this project.

This led us to design a governance model in which the network will be governed by a Foundation, probably based in Switzerland, representing the interests of different stakeholders (companies, financial institutions, technology partners) and different regions (Americas, Europe, Asia-Pacific). To remain effective, this governance will probably have to rely on sectoral associations legitimately representing the interests of these stakeholders, rather than granting each member of the network the right to vote. We will therefore have to find the right balance between sufficient efficiency on the one hand and necessary representativeness on the other.

In addition, the rules of access to and engagement in this network must be clearly established, as well as the purpose of this exchange of sensitive data, otherwise the network could quickly be turned into a purely short-term financial tool that would eventually destroy this trust capital. The Foundation's mission will therefore be to guarantee and develop a certain number of standards on data use and ethics, defined in agreement with the stakeholders.

How does the combination of this technical architecture and governance model ensure trust in this network?

It would be pretentious to say that this platform could guarantee this trust in a clear and definitive way. We would claim, more modestly, that we have designed this platform to set up the conditions so that this trust can emerge, in particular through the following dimensions:

• Trust in the identity of the persons and legal entities involved in these exchanges.

• Confidence in the technological infrastructure that underlies this network.

• Trust in the information exchange processes themselves, designed to respect the privacy of information and its authorised and consensual sharing.

• Confidence in the ability to audit and track these exchanges if necessary.

• Confidence in the governance of this network, designed to serve the Common Good, and to avoid takeover by an interest group or region.

The trust that a user will have in this platform is a necessary but insufficient element to promote its widespread adoption. As with any product, it is the application value of this service for the various players that will determine the degree of adoption of and commitment to this network. From this point of view, it is essential to consider this application value, so that this product brings real tangible benefits to its users. These benefits must be translated into concrete savings in time or money, but also into new forms of value creation associated with responsible use of the data, oriented towards trust in the business ecosystem.

Beyond the application value, it is also essential to think carefully about the business model of this platform in order to encourage its rapid dissemination. The experience of the platform leads us quite naturally to „freemium" type models, designed to limit friction during adoption.

What lessons can you draw from this complex project at this stage?

Technology, in this case blockchain technology, is never an end in itself. It is neither good nor bad, but finds its meaning only when used for a project and therefore with a purpose. It is, therefore, essential to clearly define the vision, mission and values of such a collaborative project when it is launched.

It is very striking to see the extent to which the debate on blockchain technology illustrates this point and the differences of opinion of this world view. Issues of data use and platform governance clearly illustrate this debate.

• Can sensitive private data be exchanged and used without the permission of the owner and for any purpose? Europe has answered 'no' to this corporate question and has established stringent regulations on it (GDPR). This vision is represented in our project's very architecture.

Is governance necessary beyond the technical management rules set out in the Smart Contracts of blockchain technology? In the world of business exchanges,

we thought so, in order to prevent conflicts of interest, and to limit the abuses of power that will inevitably occur. Trust in technology is confronted with the reality of the business world and its power relations, and has led us to design a safeguard that seems, to us, to supply this very human dimension of trust that is, by nature, lacking in technology.

Paradoxically, the real "safeguard" of this platform is probably trust itself.

In the business world, all exchanges are based on trust between counterparts. This "trust capital" is constructed gradually, slowly, through a brand and a reputation, as a result of the quality of its exchanges and the reliability of those involved. The characteristic of this trust capital is its fragility, and its vulnerability to any reputational or operational risk. Business leaders have grown to fear such a reputational or operational incident, because it can destroy this trust in a few seconds, thus leading to massive and sometimes fatal destruction of value for the company.

In such a network of trust, it is therefore not very rational to put this trust capital at risk. On the contrary, it would appear much more strategic to implement "best practices" and make them known within its ecosystem, in order to increase this trust capital, while remaining faithful to the corporate mission and values, which should not be faked. There is nothing worse for a brand than being perceived as incoherent, inauthentic or manipulative.

The ultimate vision of this project is therefore the implementation of a virtuous circle of trust within business ecosystems, aiming to encourage each company to build its trust capital through to its social and environmental responsibility, which probably represents its highest degree, and its strongest guarantee of sustainability.

When Artificial Intelligence goes to war

War, innovation and technology

The fields of war and technological progress are intimately linked. But before we attempt to foresee the possible implications of Artificial Intelligence in armed conflict, it is useful to consider the nature of their connection. War, as a "total social fact" resists definition through its historical and spatial universality, its extreme formal diversity, and its scale of intensity. An extreme form of political violence as a "confrontation of opposing wills using force to resolve their differences",[1] and the ultimate context of the struggle for the survival of human groups, war is, therefore, by necessity a central context for the exercise of human intelligence. Science and technology, as an expression of the effectiveness of rationalisation and human power over the environment, have found in armed conflict an essential and vital field of application.

War is thus the decisive illustration of the fact that technologies are not neutral but subservient to a political and social objective of efficiency. Efficiency as a political value underlines, in the context of war and elsewhere, the relationship between technological progress, the will-to-power, and power.

However – and this is an essential point – the linear increase in the technicality of the battlefield does not guarantee a superiority of politico-military effects. Technological superiority and its military uses are always intertwined with a political and ideological context. The mass army resulting from conscription, the birth of the modern state and nationalism are equally – if not more – crucial than the widespread use of firearms on the battlefield. The same applies to nuclear deterrence, a doctrine that depends as much on science as on political ideology and the collective imagination. In the same vein, the contemporary case of the role of disinformation in hybrid wars is not so much related to social networks and the internet as to an "innovative effector" that transforms the agents of this disinformation into a new mass army attacking democracies caught between freedom of information and these technological aspects.

Since the first Gulf War of 1991 and the subsequent "revolution in military affairs", managerial orthodoxy has invaded government instruments and security and defence policies. This de-politicisation movement, focused on the "measurability" of bureaucratic indicators and debates on budgetary quantification, ultimately weakened the political essence of the conflict in favour of a "military positivism" where technological progress would *de facto* imply efficiency gains.[2] The obsession with technical perfectionism has therefore produced "bonsai armies", too small because of the exorbitant cost of highly sophisticated equipment, and too fragile because of the complexity of weapon systems that are sometimes very vulnerable to simple weather conditions.

Moreover, the shifting and asymmetric nature of recent conflicts has demonstrated that progress in the

1. André Beaufre, "Introduction à la stratégie", *IFRI/Economica,* 1985, p. 16, cited by Joseph Henrotin, "L'innovation au sens stratégique du terme", *DSI,* Hors série No. 61, Sept.2018, p. 43.

2. Joseph Henrotin, "L'innovation au sens stratégique du terme", DSI, Hors série No. 61, Sept.2018, p. 45.

art of war cannot be reduced to technical sophistication and increased performance (faster, stronger, further). The inventiveness and "DIY" nature of armed groups have often illustrated a re-actualisation of the David and Goliath myth through a real theorisation of a "competitive techno-regression". This includes improvised explosive devices, the militarisation of commercial drones, hacking and non-electronic means of communication.

Thus, technological innovation in itself is never a source of political and military success. It is rather a question of considering it in a triptych balanced between "technology-organisation-doctrine". In short, in the dialectical context of war, its encompassing socio-political dimension and the "fog" it always generates, the key word remains "adaptation", continually "in progress", which constitutes the heart of war to be considered precisely not as "technical" but as "art".

We will focus on three areas to better understand the challenges of artificial intelligence in the military field. The tactical field of drones and other military robotics, which is very present in the public and media imagination, must not conceal that of the doctrine of use and strategy. We will therefore focus primarily on the progress of AI in the Command and Control functions of the military decision-making chain. We will continue with a reflection on the future of AI in the field of cyber security. Finally, the most promising approach within current reflections and research seems to be oriented towards a doctrine that advocates the constitution of real human-machine pairs playing on the cumulative advantages of hybrid human/non-human devices.

Artificial Intelligence and command and control (C2) functions

According to an HCSS report,[3] it was intelligence that enabled homo sapiens to reach the top of the food chain. Two types of this intelligence are invoked: the ability to create weapons for the purposes of defence and the ability to share relationships and knowledge for the mobilisation of non-physical means of warfare: propaganda or intelligence, among others. Command and Control (C2) activities are therefore intrinsically linked to our intelligence based on the definitions by Pigeau and McCann,[4] for whom command is "the creative expression of human will necessary to accomplish the mission", which therefore implies human intelligence. Meanwhile, control provides "the structures and processes to enable that expression", which is based more particularly on means. There is therefore a dialectic between these two functions, control being a tool of command. More recently, the definitions of C2 incorporate the presence of new technologies and C2 has become a scientific theoretical field of study the complexity of which is linked to the anthropogenic

3. Stephan De Spiegeleire, Matthijs Maas, Tim Sweijs. Artificial Intelligence and the future of Defense: Strategic implications for small and medium sized force providers, The Hague Center for Strategic Studies (HCSS), The Hague, 2017.
Document available on line at: https://hcss.nl/report/artificial-intelligence-and-future-defense

4. Ross Pigeau and Carol McCann, "Clarifying the Concepts of Control and of Command," Command and Control Research and Technology Symposium, New-port, RI, 29 June–1 July 1999, p 4, cited by LCL Marc Leblanc, Artificial intelligence: the future of command and control ?, PCEMI - Exercice nouveaux horizons, 2001.

Document available on line at: https://pdfs.semanticscholar.org/dbbf/53d69c80e0024a23bd054305f96ab1c88994.pdf

dimension of current command and control systems.

While technology has always been used in control functions, the impact of AI is likely to be much more revolutionary in its implementation in command activity. Since the creative aspect of the command activity presupposes that human beings are the centre of the device, AI would have to near the stage of super intelligence to be able to carry out this activity successfully. It would then, theoretically, be stronger than a human being dealing with the stress and fatigue of war situations.

In the event that super-intelligent machines are developed within the armed forces, this would mean moving from a traditional mode of confrontation, with a human being at the heart of the processes, to a war of algorithms. However, it seems difficult to imagine this type of confrontation in the medium term: the lack of experience in the field of cooperation or combat between AI does not allow us to imagine the results of such an evolution in the nature of war. More concretely, however, this would significantly reduce the chain of command, as operational decisions could be made in seconds rather than hours. This would significantly improve the effectiveness and efficiency of the armed forces. This would then raise the question of the place of humans in conflict. The notion of algorithm warfare is reminiscent of the theoretical concept of Hyperwar put forward by Marine General John Allen, and defined as "a type of conflict where human decision making is almost entirely absent from the [...] loop".

The question of automating human decision-making is central to our thinking: is it desirable for AI to replace humans in decision-making? This question in itself raises two issues: access to and reliability of information, and explainability. Indeed, since it is now impossible to clearly understand the reasoning behind machines using deep learning, it seems illusory to consider entrusting ethical decisions to AI. Finally, AI-based Command and Control activities could create a system composed of subsystems, each of which could send conflicting messages because of the mass of information provided by big data. Some people refer to this under the prism of "complexity engineering".[5]

The question of surpassing human intelligence is also crucial here. If we recall that it is thanks to intelligence that humans have been able to place themselves at the top of the food pyramid, the challenges posed by the possible development of a super intelligence equal to or superior to ours, in the field of defence, are clear. These issues relating to the loss of control and understanding of an AI used in the C2 domain ultimately raise the question of sovereignty and governability. Since the very structure of the defence organisation is inseparable from a political dimension, a profound reflection will therefore have to take place on the links between the military institution, its various personnel and its political control. Finally, with regard to international and geopolitical law, the difficulty of controlling the proliferation of this type of object, which has a very strong civil/military duality, poses another problem with regard to the political issues of sovereignty through their use in a context of asymmetric conflict between State powers and armed groups. This is particularly the case for their possible uses in the context of a cyber war.

5. Bernard Claverie and Gilles Desclaux, "C2 - command and control: un système de systèmes pour accompagner la complexité", Communication et organisation, 50 | 2016, 255-278.

Opportunities and risks of AI for cybersecurity.[6]

Artificial intelligence is already progressively present in many defence-related fields, including cybersecurity. For example, security information and event management software (Security Information and Event Management / real-time cybersecurity analysis platforms) are among the first examples of AI in this field. It is mainly in the fields of code analysis and behavioural analysis of computer systems that AI could develop new and effective methods in the fight against cybercrime and cyber-attacks. The increased use of AI, capable of processing significant amounts of information in record time, would make it possible to more effectively identify malicious codes by comparing them with code databases classified according to their nature. In terms of the analysis of computer system operations, the contribution of AI could increase the identification of abnormal behaviours that are symptomatic of attacks or security breaches.

However, several difficulties counterbalance the opportunities that AI gives to cyber defenders. First, there is a difficulty of analysing codes that results from changes in their evolution. On average, malicious codes change and evolve every two years, while AI will be slower to acquire data and know how to use it. In behavioural analysis, there is a risk associated with the ability of AI to detect certain attacks. The volume of data to be processed is so large that, despite the increased capabilities of AI, weak signals may not be identified. Hervé Debar defines these difficulties as the problem of the rule governing the political dimension of the use of information systems and that of over-investment in data.

Finally, one major difficulty of the development of AI in cybersecurity is related to the advantage that the attacker still has in this field. Due to the eternal theory of the sword and shield, the attacker's imagination should not be underestimated. The effectiveness of AI can be limited by saturation of information or by multiplying false alarms and anomalies. It will be more difficult for AI to hide its vulnerabilities than for the attacker to imagine new attacks, new malicious codes. It would even be impossible for AI to prevent attacks not foreseen in its starting code. Defensive AI may therefore always lag behind attackers' innovations. Despite the possible improvements that the development of automated AI could bring to the field of cybersecurity, there are still significant limits to this field that imply a real need to continue to keep humans in the decision-making and execution loop for efficiency and adaptability concerns.

Towards a human/non-human hybridisation[7]

The fantasy of an AI war therefore still remains largely in the realms of science fiction, not only for technical reasons but also because of the socio-political significance of an act of war. However, when we look at the direction taken by research programmes in the military field, another strategic axis emerges. Since the onset of programmes aimed at expanding the role of more or less autonomous military devices, it has undoubtedly been useful to listen to input from the military. When

6. Hervé Debar, "Intelligence artificielle, risque ou opportunité pour les cyber-défenseurs ?" Telecom, Number 190, Oct. 2018.

7. Michael Joseph Gross, "The Pentagon's Push to Program Soldiers' Brains, The Pentagon Wants to Weaponize the Brain. What Could Go Wrong?", Nov. 2018.

a skilled person is asked what they expect from a device of this type, the model that comes up is that of the "sheep dog". Autonomous enough to take initiatives to accomplish its mission, faithful, enduring, defensive and communicative with its master, as well as little animal intelligence but a relatively high degree of conceptual understanding of intentions, beyond the simple function of order or command. A system capable of adapting to the context to modify its behaviour in order to accomplish a mission of a fairly high degree of generality such as "gathering, guiding, defending a herd, or even its own master" and to generate a sufficiently empathetic relationship as between a human and an animal. Thus, military programmes are clearly oriented towards human-machine hybridisation, the machine being able to replace humans in their preferred fields (endurance, speed, robustness, perception) with increased autonomy, but with humans remaining in control of decision-making and the adaptation of means to ends, particularly political.

For several decades now, the US DARPA (Defense Advanced Research Projects Agency) has focussed its research on the symbiosis between humans and machines. The agency has implemented several experimental programmes, with varying degrees of success. But it is the development of neurotechnology, based on the construction of medical devices for interaction with the brain, that makes it possible to now consider this symbiosis: it is now possible to consider controlling a machine with the brain.

While DARPA's stated objective in the media is to restore the capabilities of wounded combatants, reality suggests that its real objective is to create human/machine interfaces that are easily applicable to the military field in the spirit of "improving" the combatant. By playing on the field of duality (repair/improvement), DARPA's game is therefore blurred. In 2012 the Agency released a video in which Jan Scheuermann, who is paralysed from the neck down, was feeding herself using a robotic arm controlled by a brain implant. A year-and-a-half later, her brain was connected to an F-35 flight simulator. There is therefore a real grey area between healing and human improvement through neurotechnology and it seems particularly illusory to imagine that technological advances leading to the appearance of improved humans with civilian applications, would not be adopted in the military field.

Improving memory through the implementation of a neural interface is another very important area of research within DARPA. A series of tests illustrated the possibility of encoding memory, and therefore learning, and transmitting it to another individual in mice. Currently, as part of the *Targeted Neuroplastic Training* programme, researchers are studying the possibility of simulating the vagal nerve to improve learning in precision shooting, surveillance and recognition, and language.

For its part, the Franco-German fighter aircraft programme of the future (SCAF - Système de Combat Aérien du Futur) has placed the concept of Man-Machine Teaming at the heart of its upstream studies. Even by 2040, the experts in this programme do not expect to be able to do without a human pilot in the field of air combat. The general idea of SCAF is enlightening in terms of the future place of AI in the military field but of course more broadly: "The principle [...] is to provide the various machine systems with more autonomy and artificial intelligence to promote an extended and reworked human-machine relationship. From this perspective, these intelligent systems would no longer be limited to the simple execution of actions requested by an operator. They would enable collaborative work that would make operators' actions and decisions more efficient and effective while saving their mental and physical resources."

Conclusion:
the future of Man-Machine teaming

An AI war is not on the immediate horizon. The technical limits of human virtues in combat (courage, adaptability, empathy, justice...) and the political scope of any act of war should encourage us all to reflect collectively on the meaning of technological progress and the assessment of its real, rather than imaginary, challenges. Such an issue could even eventually be the subject of international agreements in the wake of those that concerned nuclear energy during the Cold War.

On the other hand, the direction of military research – often historically the most decisive – towards human-machine hybridisation probably outlines the position of AI in the socio-political field. Both reassuring and promising in many areas, the question remains as to whether, by penetrating this "unknown continent" of the human brain, retro-engineering will enable illicit and negative invasions towards the human brain itself.

"War can never be identified with virtuality", Interview with Dominique Lambert

Professor of Philosophy of Science, University of Namur, Member of the Royal Academy of Belgium

Currently, a fantasy is developing in the collective imagination that involves imagining the war of the future as a dehumanised war, led by killer robots. Is this a realistic vision in your opinion?

First of all, it should be pointed out that any war leads to dehumanisation, that is, it drives humanity back to defending its basic values. But the question seems to imply "dehumanising" in another sense: that of removing the human being from the battlefield. While it is absolutely important to reduce the number of civilian and military casualties as far as possible, the idea that we could wage war without humans, through robots, is an illusion. War is a perpetual search for asymmetry, allowing us to gain the upper hand over our opponent. The loss of machines will never have the same impact as the loss of a human life. The belligerents will therefore necessarily seek to reach human beings directly or indirectly. A war without soldiers is a dangerous illusion. One day or another, if soldiers only fight through machines, it will be innocent civilians who will pay a heavy price for this search for asymmetry that can no longer be achieved with artificial intelligence, robots or cyber networks. One could speculate that any robotic, dematerialised, disembodied war will turn into an unstable situation whose outcome would almost always lead to the targeting of innocent civilians. This is why such a "victimless" war between machines should raise questions of international humanitarian law, as being potentially ruinous of the principle of distinction (combatant–non-combatant).

Would it be possible and desirable to integrate/implement ethics into the founding algorithms of AI? If so, on what basis should this ethics be based in the field of defence?

I would begin to answer by saying that we must distinguish between questions relating to the implementation of ethics in algorithms and the ethical problems raised by the very writing of algorithms (including those that are not algorithms controlling ethical parameters). Indeed, we must not allow ourselves to be fascinated solely by the question of the possibility of translating ethical and legal requirements and reasoning into a programme. Moral questions already arise in the way an algorithm is written and designed. For example, if an engineer is not aware of a certain number of biases in writing an algorithm, they can implicitly write and propagate a certain number of counter-values through the algorithm that raise ethical questions.

Being aware of the limitations and biases of algorithms is part of a programming ethic, an algorithmic ethic. Algorithmic ethics, on the other hand, consist of a claim to translate ethical standards into a program-

me. It may be useful to implement legal or ethical safeguards in defence systems. It is also very useful to have efficient automated systems to assist in ethical or legal decision-making. We must not, in fact, deprive ourselves of recourse to efficient systems that can help our discernment and decision-making. On the other hand, it may be problematic to believe that "ethical algorithms" could be sufficient for our military decision-making. Why? Because human decisions, which sometimes makes it possible to save the most desperate situations, are based on creative capacities to break out of old frameworks and invent new ways of acting. Ethical and legal algorithms are necessary. But it is important to think about a series of situations in which a human being was necessary to break out of a system of norms that made it impossible to make a decision or that forced a machine to make decisions that were contradictory to the spirit (if not the letter!) of its programme. It is, therefore, important to value everything that can help in ethical and legal decision--making, but it is equally important to ensure that programmes do not lead to situations that are contrary to the spirit of principles and laws. However, in a sense, only human beings can transgress the limits of languages to save the meaning they convey.

It should be recalled here that ethical decisions requires consideration of the purpose of the act, the context in which it is performed and the underlying intention. But taking into account the context as well as the intention requires interpretations that are rarely implementable in a formal language. As Aristotle points out in his *Nicomachean Ethics*, ethical decisions cannot be reduced to the order of a "mathematical" derivation, starting from axioms and obtaining moral theorems by rules of deduction! The application of universal norms to particular contexts requires intuition and interpretation, which would be difficult to translate into standardised, recursive and, in short, computational processes. This is the classic challenge of a reflection on the "judgement of prudence" or, in the field of law, on the links between deontic logic and legal rhetoric.

The question you are asking also involves the nature of the ethics at stake (in ethical programmes or among decision-makers themselves). It is not easy to answer this question in a world where we readily admit that there is no global ethical consensus. However, I think that we can identify common principles that refer to a requirement not to destroy humans or the environments in which they live. This requirement could be described as the "principle of anthropological non-contradiction". Respect for the dignity of persons is an aspect that comes back to the level of international bodies, and is an instantiation of this "principle".

Could AI really one day make better decisions than humans if faced with a military ethical dilemma?

Often in dilemmas machines do not do better than humans. Indeed, if humans knew how to deal with dilemmas systematically, they would have found a way to programme machines to do this efficiently and quickly. The specificity of the dilemmas is that there is often no set rule for getting out of them. But a decision has to be made by "inventing" a possible way out. This is sometimes done by accepting a personal sacrifice due to higher values. What makes the crucial difference between the human and the machine in the case of a dilemma is that the former will take a risk and accept full responsibility by choosing a solution. Accepting responsibility means that the individual agrees to be held accountable for their actions and possibly

to pay the price for their decision. The answer to your question is, therefore, as follows: the machine may sometimes be able to find a solution that will make it possible to get out of an apparently inextricable situation (because it has the ability to explore all solutions faster than humans), but in cases where there is no optimal solution or in those for which all solutions lead to disasters or tragedies of the same intensity, only humans can take responsibility for an inevitable action and pay the price. It should also be noted that in some cases, humans can, through their ability to think "outside the box", find non-standard loopholes that can save the situation. Once again, somewhere we have to reserve a place for a human decision-maker with their values and sense of responsibility.

What would an algorithmic war look like? Would it lead to the dehumanisation and de-politicisation of war? If so, is this desirable or would it lead to more and more asymmetric wars?

Algorithmic wars already exist in intentional actions of hacking and disinformation. I wouldn't talk about this as a dematerialisation of war, as this is not very fair because cyber war will lead to consequences with terrible material effects (hacking into an energy plant can lead to deaths in hospitals, etc.)

The question of algorithmic warfare raises similar questions to those raised by the use of social networks. The consequences of some statements may seem harmless because they are written on an individual screen, but when you see the kind of very concrete reactions they provoke, you grasp the harmful power of broadcasting certain news, fake news, etc.

This blurs the very definition of war and is one of the most crucial problems of cyber warfare. How can an algorithmic state of war be correctly and legally defined? One could say that this type of cyber warfare will make a new field of intrusion possible. We must get used to the idea that war is no longer fought only in geographical theatres. War now infiltrates cyberspace. However, cyberspace is part of the individual and social space. Just as the territory of some countries extended at one point into territorial waters, now we must see individual and national spaces extending into the virtual dimension of cyberspace. And it is in this dimension that new conflicts, intrusions and attacks can emerge.

There is another dimension to the question asked. Algorithmic warfare also means the authorities' increasingly systematic use of military decision support systems. It may indeed be that war games, operational simulations, etc., lead politicians to do nothing more than follow what the machines suggest. In this sense, this could lead to a form of de-politicisation.

Another risk I can see in algorithmic wars, in cyber--conflicts, is the problem of acceleration, of runaway systems. I am thinking of an escalation of violence linked to ultra-fast response systems. This type of rush can lead to defensive responses that are no longer proportionate and therefore no longer compatible with one of the important principles of international humanitarian law. The search for asymmetry is part of conflict, but this desire to achieve victorious asymmetry will lead to escalation and disproportionate responses through the speed and power of the processes. Perhaps it is worth recalling that the very fact of talking about war between algorithms is ideological! War is always dehumanising and always ultimately affects human beings. To speak of "war between algo-

rithms" obscures the fact that they inevitably lead to harmful actions outside cyberspace, in places where potentially innocent beings end up paying the heavy price of violence. It also obscures the fact that those who start and fight wars are never machines but human beings. It may be worth noting that while artificial intelligence can contribute to war, war is never "artificial", it translates into real and well-embodied physical or psychological effects in real people.

Military research seems to be moving towards man-machine hybridisation, what does this medium-term approach taken by military programmes mean to you?

If a machine can help to make decisions that promote peace and safeguard the dignity of people, then man--machine hybridisation makes sense. But we must be careful, because humans are fascinated by the efficiency of machines and may tend to systematically abandon important elements of their powers in their "mechanisms". If hybridisation occurs through irreversible alteration of human capacities or organs, this raises enormous ethical questions. Some hybridisations could indeed be designed along the lines of implants that are irreversibly implanted into soldiers' bodies. This type of practice would harm the soldier's physical integrity and they could very well become, when demobilised, disabled (leading to a new type of war wounded).

But even if we are not thinking about this kind of "physiological" hybridisation, important questions can already be raised about the responsibility of people who are immersed in networks of machines or systems that help them and push them to make decisions. What is problematic is a kind of crumbling of responsibilities and leaders. In the event of collateral damage, there may be problems in identifying those responsible if decision-making is linked to a complex network of human and machine actors. This screening and dilution could be used to more easily cover reprehensible actions.

In a man-machine system, it seems to me that the system must remain at the service of the purposes prescribed by the human being and that responsibilities must be clearly identifiable. In the case of complex systems this can be very difficult (this difficulty is already present if we think about proving that a complex system will be reliable and will respect the prescribed purposes in all configurations of use).

To conclude this interview, I would like to say that war can never be identified with virtuality, with artificiality. The risk of cyber warfare, of these expressions talking about wars of algorithms, is to remove from their horrible content a reality that always begins with human ideas and intentions and ends with consequences for the minds, hearts and bodies of flesh-and-blood people. Artificial intelligence can be used for the purposes of defence and peace, that is undeniable, but we must be wary of the screen of artificiality which could give the impression of war without consequences and without responsibility or liability. There is a need for a programming ethic that assesses the biases, limitations and implicit intentions of algorithm content. But we also need an ethical approach to the use of artificial intelligence in the military field, which does not lose sight of the reality of the effects of artificiality, in the name of defending the dignity of individuals!


UNIVERSITY
OXFO
E, IDEAS AND TECHNOLOG
HE UNIVERSITY OF OXFORD
www.innovation.ox.ac.uk

IS THERE A JUST CYBER-WAR?

Based on an interview with Dr. Mariarosaria TADDEO

Researcher at the Oxford Internet Institute and Deputy Director of the Digital Ethics Lab, University of Oxford, UK.

War is governed by a set of rules and rights, even in a context as chaotic as armed conflict. How far do the categories of the ethics of armed conflict apply to cyber-war? To what extent can cyber-conflict be considered an act of war? War. You may hear of war on a daily basis in the news, trade war at least, the escalating battle between the US and China by means of tariff offensives. Perhaps less apparent but important nevertheless is the topic of cyber-war. It is high time for a more fundamental debate to frame the question of cyber-war in a tangible and comprehensive terms. Such is the appeal of Mariarosaria Taddeo, a prolific researcher on the ethics of cyber-conflict at the Oxford Internet Institute (OII).

This article introduces the theme of cyber-conflict and its particularities. By doing so, it aims at setting the foundation to initiate a debate and involve the necessary actors at different levels. Operating in a grey area for about ten years and fueled by sci-fi phantasmagoria, the situation should not be left out anymore. It could even go further and argue that there is an urgent need to define a set of rules, right and values, which is only possible by truly understanding the potential damage of digital attacks. When establishing a legal framework, it has to mitigate risk without slowing down the amazing potential for technological development. Ethical issues are not easy to address, more over if we want the issue to be solved rapidly.

First thing first, we need to understand what attacking in the cyberspace means, and what terms we should use. Referring to it with the term "cyber-war" serve to underline the gravity of the phenomenon. However, it would be more accurate to talk of "cyber-warfare", translating the idea of conflict. "War" is, by definition, an aggressive violent action against the territory of a state performed by another state, and this action is preceded by a declaration, or an aggressive attack. We have regulation for war.

When it comes to "warfare", so to conflict, it is not preceded by a declaration of intent or followed by invasion or retaliation. No matter how evident, how aggressive the attack is, it would not be considered as war since you don't have the crucial element of intend, the declaration of wanting to engage in a war against another state. It's kind of a "softer" war of dealing with international relations. But still, there are regulations, so they have to be respected. That being said, tensions might eventually escalate to the point of declaring war.

Emergence of a New Phenomenon

The problem is, the framework that we have for the past more than two thousand years since we have thought about conflict and war, always had to do with violence and physical damage. A tangible reality, with physical causality, visible destruction and palpable suffering. Now we have a phenomenon which is not causing physical damage, with the exception of only one case, but it is still very violent, it is still very aggressive, and very effective. When it comes to the digital, the legal framework no lon-

er seems appropriate. Where to determine the limit of the intolerable if we were to compare a virtual invasion to a physical one? How to identify the enemy of a cyber-attack, in order to retaliate? And what would be a proportional retaliation? Which targets are to be saved?

Cyber-conflicts can be very violent, very aggressive, and dreadfully effective. A virtual damage can actually cause real losses and real suffering for society as a whole. Therefore, the principles of proportionality, necessity and discrimination, as well as the moral values derived from the doctrine of just war are to be transposed into the cybernetic world.

Establish an ethical framework for cyber-conflicts requires in-depth reflection. Taddeo and her colleagues often remind us that a transposition of regulation in place applied by analogy is not enough. They suggest rather to translate the ethical justifications (under which conditions to conduct a conflict) and the form (which line to adopt) of this new phenomenon in its virtual context. We have to understand that cyber-space is a different deal, one that has its own particularities, its own dynamics where technology allows doing things new, and with things, so the regulation has to be designed on the basis of the profound understanding of these technologies and how it behaves. In the end, the goal is to guarantee a peaceful coexistence, in the same line as the just war theories do for the kinetic world.

The War Effort

Ethical justifications are necessary to enhance legitimation of violence for retaliation, defense and preserving global order. In a way, it mobilizes public support. In virtual space, it is not easy to awaken a sense of fraternity and collaboration to the collective war effort. The evidence is not so under the sky, let's say. You know, if you bomb the place you will know that the place has been bombed. But if you launch a cyber-attack, public opinion may not know about it. The problem is that our society has been increasingly more reliance and dependent on information infrastructure, hence more and more exposed and vulnerable.

The case of Estonia is revealing in this respect. We often reference to the 2007 cyber-attacks by denial of service as the first "cyber-war" between Estonia and Russia (as defined above, we still think that conflict would be the adequate term here). The population, informed in full transparency by its government, participated to a coordinate effort, from volunteering to education and white hacking. The citizens' relationship to the digital, as well as the trust in their government, are permanently transformed.

Transparency cannot be taken for granted in a cyber-attack. We have to find a balance. On the one hand, confessing to being a cyber-victim could cause panic, harm the credibility of the system put in place and undermine trust in the government responsible for protecting its population. On the other hand, silencing an attack allows hiding one's failures from other potential enemies, contain the situation, and even retaliate without being held accountable.

Nuanced Regulation

Cyber-conflicts have been going on for a while now, more than fifteen years, and it has created a new kind of opportunity for state actors to run their international relations. It is very easy to use these tools behind the scenes, and it is to some extent advantageous to leave them unregulated or underregulated, just be-

cause everything gets sorted in a kind of tacit way. In the same line, there is the way of using cyber-means for espionage for example, which tend to be hidden. There are different measures, at different levels on which we must focus on. It is nuanced, and regulation must take it into account.

Everybody that deals with cyber knows any cyber-system, any information system is never secure 100%. That's what the nature of the system is. As soon as a system is created, there are also ways to attack the system to breach the system. The pervasiveness of these systems, including for key national infrastructure, increases our vulnerability to cyber-assaults. If a cyber-attack is able to target a critical national infrastructure, which is what we fear these days, there is no chance that the cyber-led attack can be hidden. Because the knowledge would be too big. The damage would be too big for the state not to make a case of out it in order to retaliate and respond to the attack.

Should we wait until such a threat materializes to legislate? Well, I hope that it doesn't take a crisis. There is a pressing need to regulate state behaviors, this is of crucial importance, and it is too bad that the serious attempt by the UN was doing failed a couple of years ago. We cannot postpone it any further. Several states have started individually to define measures of cyber defense, which are called active cyber defense. The pressure to define an ethical framework should come from single states, member states of the UN, member states of the European Union, member states of the NATO. We underline how crucial the international level is in trying to regulate a new phenomenon, something that we have never seen before. We need to understand what values are there, what values we want, what principles we want to embrace. We need to create a trusted environment to allow for a high-level dialogue. It needs to be constant. It also needs to be transparent, in terms of who's in and who's out. Finally, it needs to move both horizontally and vertically.

A Constellation of Actors, Horizontally and Vertically

Throughout the ages, it has taken the synergy of philosophers, theologians, ethicists, military leaders, political decision makers and lawyers from all over the world to develop a set of conventions and agreements to guide the act of war. This horizontal synergy must now include private sector expertise in the high-level dialogue. We know most of the skills, the highest skill in the term of designing information-system or cyber-system, or in terms to attacking them and in terms to defending them, they reside within the private sector. Engineers can help us understanding technologies, but regulations reside in the hands of the state. Same as for weapon producers, it would be tantamount to give them any duties, or even any chance to contribute to the regulation in these areas. It is too big a stake.

The highest authorities must set the first regulatory milestones. Time is running out, we need regulations, to make sure that states are accountable for what they do in cyber-space, at least as much as they are held accountable in the way they would be behaving in the physical domain. And this regulating needs to be aware of the importance of shadows, technologically speaking the affordance costs of these technologies. Because many actors are acquiring new skills, aggressive skills and capabilities, encouraging a kind of cyber-arm race, even an escalation towards armed conflict.
For that you need to have an authority, like the NATO or the UN, to be designated to assume the role of mo-

nitoring, redressing and enforcing sanctions. It must be politically costly for States to derogate from the laws of cyberspace.

One of the examples is the disclosure of vulnerabilities. Vulnerabilities are basically glitches in the system, which once identified can be corrected. Taddeo compares it to an open window in your house. Maybe nobody notices to enter in your house. A disclosure of vulnerabilities would be like a neighbor noting the open window and warning that there is a risk of burglary. Not warning should be considered as an illegal behavior in cyber-space. This is the case of WannaCry, which abusing from a vulnerability to entering into Microsoft system. The NSA, the National Security Agency of the US saw this vulnerability and they didn't disclose it to Microsoft, so that Microsoft could patch it, because in that way they could abuse it to entering into the system if they wanted to for security or espionage, for example. This should be a behavior that once it becomes evident is sanctioned by the UN for example, or any international law. Because it endangers the security of a lot of states, even endanger the security of a lot of individuals.

It is not only about sanctioning, but also about improving collaboration with a trusted environment. in which allies work together to define and improve their skills. For example, NATO has treaties with private industries to improve and understand the cyber-means and cyber-technologies. It also organizes a kind of cyber-exercise. It is not mandatory yet, but imagine the benefits if it would. In addition to the benefits not specific to the virtual world (strengthening alliances, consolidating defenses, improving strategies, sharing expertise technical, etc.), these training sessions would make it possible to test and refine AI-based cyber weapons thanks to the influx of data. Artificial intelligence improves by being used, the more you use it, the better it becomes.

Collective effort does not stop at the decision-making level. Discussion could be vertically as well, in the sense of going down from higher level to some lower level. We need to educate people to what cyber security means. The WannaCry attack, which was kind of a massive phishing attack, was successful because of the lack of understanding from the lay public. In any security system, people are often the weak link.

Education vs. Entertainment

Taddeo warns that science fiction cannot be mistaken for education. It's a distraction. In addition to being potentially anxiety-provoking, science fiction is getting problematic because it distracts us from this debate. For many, it's still a debate, the debate of singularity, the debate on the idea that the machine will become so intelligent, so powerful, so evil, you know destroy us all or enslave us all.

This kind of AI as a magic evil is not scientifically grounded discussion. Artificial intelligence is nothing else but a computer which is able to perform actions in a way that if a human performed the action you would say that the human is intelligent. But artificial intelligence has nothing to do with intelligence: they don't have intuition, they don't have ideas, they don't have emotion, they don't have cognitive processes. The way of putting it is that the problem is the way we use it, it is not about what AI does or does not. This does not prevent this artificial intelligence from being autonomous in the actions to be performed: selecting a target, launching a virtual attack in the system, retaliation, etc. This form of.performative intelligence will not consider the consequence of his actions. And for now, neither do we. The problem of legislation lies in the way we deploy these machines: they don't wage conflicts; they

are tools that we use in such a way that we can wage conflicts. They must be held responsible for the machines deployed, and be punished if necessary.

AI: Tool and Opportunities

At this stage, let's focus on the skills of cyber-weapons, especially those based on artificial intelligence. Cyber-attacks require relatively few resources for a global reach. They are anonymous, decentralized, and operate on a system that is by nature interconnected and porous.

Infallible defense is technically impossible, any barrier can be overcome with time and sophistication. Therefore, the best strategy is active cyber-defense. This is where artificial intelligence comes in, as an active defense tool. By monitoring activities, detecting recurring patterns, tracking anonymous attacks, algorithms could try to automatically track back the attackers, and even try to retaliate the attack. We are only beginning to use these AI capabilities, but it is clear that without a legislative framework, these tools will make the Internet a constant battleground.

AI Is Not a New Enemy

The weaponization of artificial intelligence and the potential increased surveillance of behavior on the web are two threats to the near future of technological development, as Taddeo acknowledges. But the consciousness-raising message she carries is that the biggest risk is if we don't see this good and we don't take advantage of it. It would be a huge missed opportunity if we did that. She repeats, AI is not the enemy. First, because there is always a human leading behind. Second, because these risks can be mitigated and resolved with the right policy and the right regulations.

Because of the fears, we might not be able to make ethical decision or resolve mass surveillance. Fears may lead to a backlash. And this would be a huge mistake. We might just get into a let-die thing and miss out the opportunities that AI is brining. Governments and international institutions should not, for fear of eroding confidence in the economy and political institutions, impose a brake on the digitization. To avoid the militarization of cyberspace, let's instead create an ethical framework that allows technologies to flourish in our society.

Strength Comes From Unity

Chinese wisdom says it all: it is in [potential] danger that the opportunities lie. Speaking of China, the country understood the importance of seeing the good in AI and getting the most out of it. The government announced "A Next Generation Artificial Intelligence Development Plan" 2030, a national strategic plan that it is all centered in that fusion between civilian and military developing this technology. The use of AI in the military is not seen as completely independent or different from the use in civil society.

It is a technology that is extremely malleable. This approach allows costs to be very spread across sectors and society, while fostering progress. GPS, the Internet, drones were all born as a military project. Their application in other contexts greatly benefits economic development and steadily improved our quality of life.

"Peace" in the military sense of the word is synonymous with latent war, said the American philosopher William James. Cyber-peace, too, is not exempt from preparation, attacks and counter-attacks. The doctrine of just war must be translated into the digital age, but the background remains unchanged: then as now, it legitimizes the defense and measures used to maintain peace and create space of successful individual and collective development.

Rapport ETHICS & TECH 2019

Part IV
INNOVATION AND ETHICS *BY DESIGN*

In this last part, we will explore the relationship between ethics and innovation in light of contemporary phenomena which are often overlooked, such as the development of China's huge makerspaces. We will also consider reflections on the compatibility between innovation, growth and sustainable development, frugal technologies and the latest views on the legal ethics of AI.

As is often the case, passing media trends mean that the reality of technological innovation phenomena is seen through a qualitative and quantitative filter.

This distorting prism produces a narrative of innovation represented by mythical figures, fantasised places and utopian futures. Of the various ideologies at work, that relating to disruption is particularly striking.

However, in recent months, this heroic tale has given way to a growing concern about the genuine social utility of these vast movements of supposedly creative destruction.

Positive externalities are slow to emerge and, week by week, disruptive success story risks turning, in an equally hypocritical way, against the organisations that were at the pinnacle of innovation a few years ago.

The authors who contribute towards the following pages all underline the urgent need to work collectively towards a technological democratisation capable of demystifying oligarchical ideologies, of placing trust in the collective creativity of users, and of steering the adoption and regulation of technologies towards the universal concept of the Common Good.

créativité collective des usagers et d'orienter l'adoption et la régulation des technologies vers le concept universel de Bien commun.

Contributors:

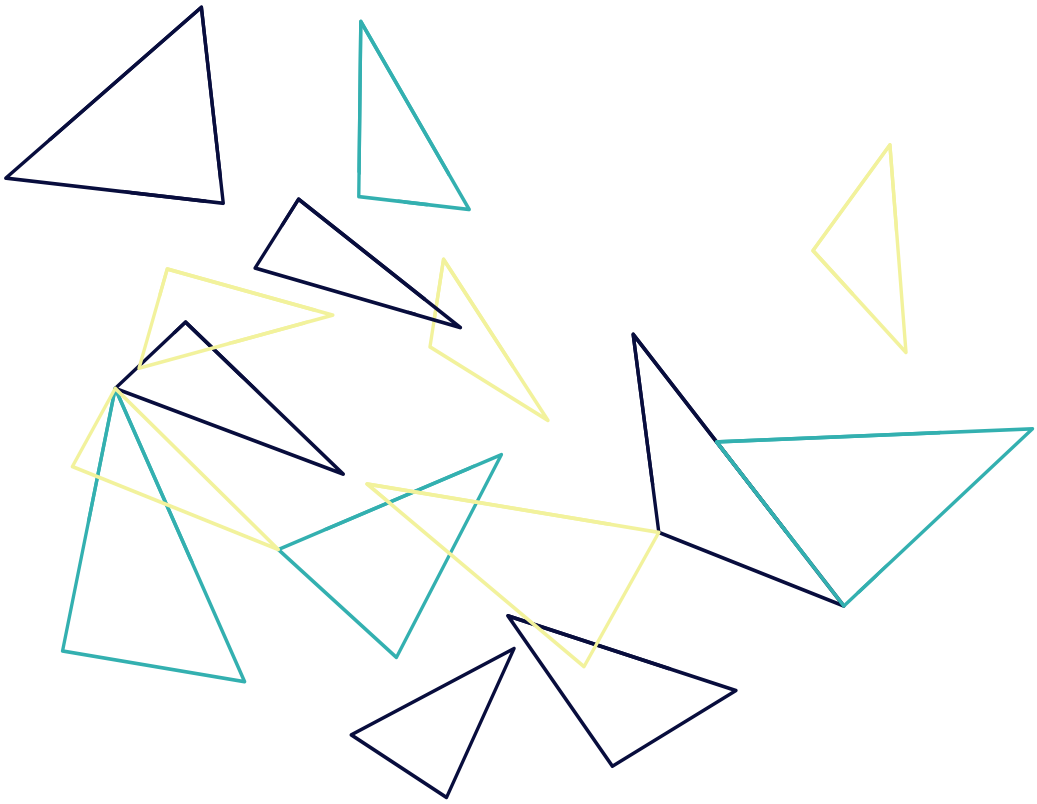
Marylaure **BLOCH**

Joanna **BRYSON**

Bogomil **KOLBRENNER**

David **LI**

Bertrand **PICCARD**



CONTENTS

Disrupt Together

David LI, co-founder and director of Hacked Matter, Maker Collider and Shenzhen Open Innovation Lab. He is a pioneer of open-source in China with the founding in 2010 of XinCheJian, the first Chinese makerspace

"Many small people, in small places, doing small things, can change the world."
(Eduardo Galeano)

The year of 2018 was a turning pointing for the digital economies. CEOs of large digital ads companies such as Facebook and Google were summoned to testify in front of the US Congress over their business practices in amassing large amounts of personal data and exploiting the data for business gain in what is termed "Surveillance Capitalism" coined by Shoshana Zuboff. In Europe, GDPR went into effect as the EU regulators attempt to rein in the influences and controls of the digital giants. The public began to be aware of the negative consequences of the digital business.

At the same time, the stories of Artificial Intelligence, many of which came from the press releases of the AI industries lead by the digital giants, were popular in the media and promoted a techno-determinism future that the AIs would obsolete human shortly. The stories of all-powerful AI under the seemingly unstoppable digital giants paint a glooming future for the humanities. Governments, civil societies, and academies around the world convened conferences and meetings to discuss how to rein in the power and influence of these giants corporations. The discussions often center around a techno-deterministic dystopian that neither governments nor societies could do much to change its course. The future seems to belong to these large digital giants and humanity is doomed without much challenge.

The digital giants are not infallible. At the end of 2018, Apple CEO Tim Cook announced that Apple missed the iPhone sales target in 2018 and cut the forecast of iPhone sales for 2019 citing the slowdown of the growth of smartphone. However, that was not the whole story. Huawei recorded a 20% growth and overtook Apple as the second largest smartphone brand in the world. Vivo, Oppo, and Xiaomi all had record business growth in the same year, mainly due to the rapid growth of markets like India and Africa. Along with several hundred smartphones brands, the mobile industry of Shenzhen took more than 70% of global smartphones market. The smartphones markets no longer dominated by a few large brands but distributed over a few hundreds of brands that collaborate in an open and sharing environment.

The rise of Shenzhen technology ecosystem defies the narrative of "creative destruction" in which breakthrough technologies originated from outside to bring drastic change to the order of the industry. "Mass Flourishing" model proposed by Edmund Phelps in which the technologies transfer and diffusion meet

highly dynamic societies and the combination empowers mass participation and indigenous innovations in the creation and productions of new goods.
The reform and opening of China four decade ago, the cluster of fifteen fishing villages in the Pearl River Delta has bloomed into the modern and innovative metropolitan of Shenzhen. The city embraced the technologies advancements with a high level of dynamism that leads to mass participations turning the technologies into useful products in an open and collaborative ecosystem full of indigenous innovations. The open system eventually overcomes the dominance of a few and become the dominant force in the development of new digital technologies.

This article drew on the researches of Shenzhen open innovation ecosystem and examined its development through the theory of Mass Flourishing to present an alternative future of a distributed and collaborative innovation system that could rein in the technologies to serve the societies and communities, rather than the interests of a few giants. The open technology ecosystem coupled with the new economic model such as platform cooperativism would enable the power, opportunities, and benefits of new technologies to be more evenly distributed.

Legacy of Schumpeter's Gale

Schumpeter's gale also known as the "Creative Destruction" is the root of the narrative for modern innovation. In theory, scientists and inventors outside of the industry would discover and invent new technologies. Working with visionary entrepreneurs and insightful financiers, they eventually lead to the destruction of the existing order and wealth of the current industry and give birth to the new one. While Schumpeter's analysis is far more complex, the term has been taken on face value as Tweet size rhetoric to suggest that all things of new technologies could disrupt and destroy existing business order.

The third industrial revolution revolves around the development of ICT (Information and Communication Technology), particularly in Silicon Valley. With the success of personal computer and the Internet, the young genius working out of garages inspired to challenge the status quo and "change the world" has become the role models, and their creations are the forces of "creative destruction." This meta-narrative of young tech wizards bringing technologies change to reshape the world has resulted in the culture of techno solutionism with the motto of "move fast and break things" so they can play fast and loose with regulations and societal norms in the name of advancement by technologies development.

In the past few years, the world began to take notice of the dire consequences of the toxic culture on the economies and democracies. The societies put the business practices of the digital giants under the microscope, and the governments began to make inquiries and set up regulations.

Mass Flourishing

"Mass Flourishing" is the work of the Nobel Laureate Edmund Phelps and provide an alternative theory of innovations to the "Creative Destruction." Phelps proposes the technology transfer could initiate the economic growth from outside, received by the highly dynamic community to take advantages of the technologies. Once accepted, the knowledge of the new technologies are quickly diffused into the community and start to generate incremental and innovations for new products and new methods of productions. Next, indigenous innovations emerge to support sustainable economic development.

The smartphone market of 2018 illustrates how "Mass Flourishing" has transformed Shenzhen in the past few decades. While Apple posted a warning on its earring in 2018 and lowered the forecast of iPhone in 2019, the mobile brands from Shenzhen enjoyed record growth and took over 70% of the global markets. The Shenzhen brands are especially popular in the rapidly growing emerging markets such as India and Africa. The 70% market share of the mobiles phones are not held by a few brands but composed of hundreds of brands in a variety of scales. The development of Shenzhen and its open mobile ecosystem does not follow the theory of creative destruction. The titanic shifts in the mobile phone markets brought about by Shenzhen mobile phone industry was not caused new breakthrough technologies but by rapid technologies transfer through the open sharing and collaborative ecosystem that enable the mass participation in the business with indigenous innovations to continue to improve products and methods of productions in a rapid speed. The city of Shenzhen has become the leading innovation hub termed "Silicon Valley with Hardware."

A Short History of Shenzhen

The worlds

The Shenzhen is one of best examples of how the democratization of knowledge and technologies and commoditization of production can rapidly transform a region from the collections of 15 fishing villages of 300,000 people to a mega-metropolis of 15 million responsible for 90% of the electronics productions.

Shenzhen came into the global spotlight in 2010 with the workers committing suicides in the Foxconn factories making iPhone. The city was seen as the global sweatshop. Within short eight years, the city is now

known as one of the leading global innovation hub termed "Silicon Valley of Hardware."

The city of Shenzhen became the first "Special Economic Zone" in May 1980 as China began the "Reform and Opening" and seek to experiment with entrepreneurs driven capitalism. The opening of the city started to attract the original equipment manufacturers (OEM) of ICT devices in Taiwan, Korea, Singapore, and Japan to set up productions here. The Asian countries have been the destination of the outsourced manufacturing for decades, and the rapid expansion and relocations into Shenzhen has driven the initial growth of the cities. The transfer of knowledge and technologies came along with productions. By the 90s, Shenzhen had hundreds of thousands of factories and hundreds of technical solutions houses providing engineering and production support to the factories.
DVDs players that read every disc.

The new middle classes with disposable income emerged in China in the late 90s, and they wanted entertainment. The most popular entertainment of the time was DVDs and VCDs. DVDs and VCDs of pirated contents were popular, but DVD players from large brands had difficulty reading those pirated DVDs. With the engineering and productions capacity of ICT devices in place, companies in Shenzhen created the DVD players that reads every disc, and it was an instant hit in the Chinese market and later on other emerging markets in South East Asia, South America, India, and other emerging markets.

The massive demand of the DVD players brought substantial business opportunities to Shenzhen at the time of weak IP protections. The factories were busy producing to fulfill the market demands while the technical solutions house quickly evolved the engineering of the players in kind of forced open and sharing fashion by providing the turnkey solution for the productions. The technical solution houses shared the design of the players, somehow forced due to the weak IP protections The sharing practice evolved into Gongban (roughly translated public circuit boards) whose plans are widely available to the solution houses allowing rapid iterations of the system to improve performances, reduce costs and add new features.

The success of DVD players kicked the collaborative ecosystem with two significant features: First, open sharing of the design and engineering and second, leaving the open and wide availability of the open components, the producers could serve market needs in all regions.

The growing business of DVD also set the foundations of the open ecosystem in Shenzhen where the traditional vertical integration of design, engineering, and production within one company broken down to the multiple independents and collaborative units of industrial design, technical solutions and factories.

Followed in the DVDs, Shenzhen started to pump out more local creations of MP3 and MP4 until the next golden opportunities of mobile phones in early 2000.

Shanzhai Mobile: Folk Art of Shenzhen.

While many are sticker shock with the latest iPhone cost around $1,000, it helps to remember that a Nokia GSM phone cost around $800 without contract back in 2000. With the GSM networks complete its deployment in developing regions like China, the demands

for inexpensive feature phones started to rise. While the major brands still concentrated their effort on the developed areas such as North America and Western Europe, Shenzhen ecosystem began to make inexpensive mobile phones serving the markets not attended by the major brands. The design studios, technical solution houses, and factories complex in Shenzhen got into action. They mix and match different looks and solutions to produce phones for the developing regions such as third and fourth-tier cities and rural area of China.

Initially, the designs were hugely inspired by major brands such as Nokia and Samsung and were marketed under funny names such as Nakia or Somsong. Affordable to the developing regions, the Shanzhai phones took off, and Shenzhen ecosystem started its first gold rush making feature phones for emerging markets ignored by the major brands. Shenzhen would like soon to ship tens of millions of Shanzhai phones annually.

Shanzhai phone industries while proliferating were still looked down by the major provides of chipsets for mobile phones. The major vendors such as Broadcom and Qualcomm required large upfront license fees to access their technologies, require vast and long engineering effort on the vendor to adopt their chips for production and deal only with major brands. MediaTek, a small Taiwanese supplier of chipset for DVD and MP3, started its mobile product division in 2004 to provide the turnkey solution to serve the Shanzhai markets allowing inexperienced vendors to quickly develop new feature phones with a small upfront licensing cost and little engineering efforts. The Shanzhai flourished with mass participation in making feature phones for all sort of niche markets around the world.

While most people thought of Shanzhai as nothing more than Nokia knockoff, the reality of Shanzhai by late 2000 transpired filled with all sort of new feature phones. One of the examples was Thunderstorm phone with seven speakers that could play music as loud as a boombox. The phone was created exclusively for the workers in construction sites. They need some musical entertainments but could not wear headphones that might prevent them from hearing the warning of dangers on site.

The Shanzhai open ecosystem drastically lowered the barrier to create new phones, and in turn, the companies could afford to develop new features to address small niche markets. Shanzhai democratizes and commoditize the creations of new phones. "Shanzhai: the folk art of Shenzhen" coined by a collaborator while touring the Huaqiang Bei electronic markets, is probably the best description of Shenzhen's relationship to mobile phones. While the rest of the world still think of making a mobile phone as a high tech venture, companies in Shenzhen are mixing and matching designs and solutions to generate new mobile phones to fulfill new niches. They act on instinct, validate the phones by directly test them with consumers and iterate fast for improvement. The practice is very much like small villages making folk art rather than high tech companies engineering new phones.

Surround the cities by the villages

One of the most famous strategies used by Chairman Mao to take over China was "surround the cities by the villages" by aligning with the needs of the bottom of the societies. The smartphone is the most obvious example of how this strategy and the influence of Shenzhen in the shaking up the global industries.

In Q2 2018, the top 5 smartphone companies are Samsung, Huawei, Apple, Xiaomi and Oppo. Three of the companies are from China, and two are from Shenzhen. Xiaomi could not have existed without the supply chain of Shenzhen. In total, Chinese vendors are now almost 70% of the global market share. The Chinese vendors did not take on this lion shares by taking on the dominating brands of Apple and Samsung head on but by serving the markets ignored by the large one. Over time, the technologies lift all, and the demolishing of return on the technologies kicked in for the Chinese brands to catch up and grab a lion share of the market. In the two most important emerging markets of the smartphone: India and Africa, the competitions are among the Chinese brands leaving two major global brands behind. Tecno from Shenzhen with 37% of the market dominates African market. The share of Chinese smartphone went from almost not existing to 70% within a decade of the introduction of the smartphone.

The open ecosystem did not overtake the market by cutting edge technologies but by the open, collaborative ecosystem that allows furious competitions by many to quickly iterate and adopt the technologies to serve niche markets. Over the long term, the open ecosystem wins over the proprietary models.

With the production of 90% ICT products for the world, Shenzhen's open ecosystem is playing a now critical role in the future of creation, development, and production. Understand how the system has affected the existing mobile devices market will provide an insight into the how the future of IoTs will develop.

Development of Digital Computer Revised

Many attribute the success of Shenzhen to be Chinese specific and try to tie the open and collaborative practices to the Chinese culture. However, examining the history of the digital computer through the lens of "Mass Flourishing" would show the development of digital computers was also driven by the same technologies diffusion meeting the high dynamism pattern. Furthermore, the diffusion pattern of computer technologies also exhibits exponential growth in term of numbers of people with access to them.

In the beginning in World Wall II, the digital computer was created along with the technologies in the secret world of code breaking. There were only a handful of people in the world had access, and the production of machines took years and cost millions. Post-WWII, the digital computers were exclusively in the domains of militaries, governments, and large corporations. "I think there is a world market for maybe five computers," a famous quote attributed to Thomas Watson, president of IBM, 1943. The commercialization of transistors and the subsequent microprocessor by companies like Fairchild and Intel in Northern California in the 60s opened up the opportunities to young people like Steve Wozniak and Steve Jobs to create a cheaper version of the computers. The PC was not widely available until the incremental innovators like IBM, Compaq and later on Dell and others working out the incremental improvements to start making personal computer more affordable. Subsequently, Taiwanese OEM in the 80s continued to make incremental improvements to PCs to bring them available to the mass. The mass markets of personal computers were not created by a stroke of genius, and the

inevitability of technology took its inevitable courses to reach the mass. It was the participation of large numbers of people and companies to continuously improve the technologies incrementally to bring the personal computer to the public. In every stage, the knowledge spread, the technologies diffused and the production cost dropped. The number of people with access to the ICT technologies grew exponentially over the years to bring about the third industrial revolution. The technology enabled but high dynamic people made it work.

The Maker Movement

With the publication of "Makers" by Chris Anderson in the mid-2000s, the Makers and the Maker Movement has generated vast interests all around the world. The Makers was seen as a paradigm shift in innovations and the Next Big things were expected to come out of makerspaces around the world following the same narrative of "Creative Destruction." Cities around the world support the make spaces expecting little Silicon Valleys to grow out of them. However, the makerspaces had not yet delivered after a decade.

An alternative of the Maker Movement is as the continuum of ICT technologies diffusion originated in Silicon Valley in the 70s, spread to Asia through outsourcing in the 80s and brought to Shenzhen in the 90s. The Internet makes the spread of knowledge and the sharing of projects readily and abundantly available to many around the world. Coupled with the rapid cost drop of open source hardware for ICT and digital production, the Maker Movement became the global network of continuous ICT technologies diffusions to the mass.

Innovations are driven by access to resources: knowledge, technologies, productions, and funding. Traditionally, the accesses to these resources are exclusive to the few. Over time, the knowledge was transferred, the technologies diffused and production commoditized. The access was broadened to more significant numbers of people over time at an exponential rate as the world got connected by more and more innovations in areas such as transportation and communications. With more people having access to these resources, the speed of changes increases drastically at almost an exponential rate.

The history of the digital computer often uses Moore's Law with its exponential growth curve to promote the ideas of the inevitability of the technology. However, we should examine closely from the human side. The growing number of people who get to innovate with the technologies also grow exponentially. More people have the opportunities to innovate with the technologies, more applications are discovered which lead to more people involved in the innovations with the technologies. The exponential growth is less about the technology, but more about the number of people get to innovate with it. As the digital technologies continue to take the shape of mobile phones and smartphones, the people factors have reached a critical mass in Shenzhen that leads to the dominance of the Chinese phone vendors in the mobile phones and smartphones markets.

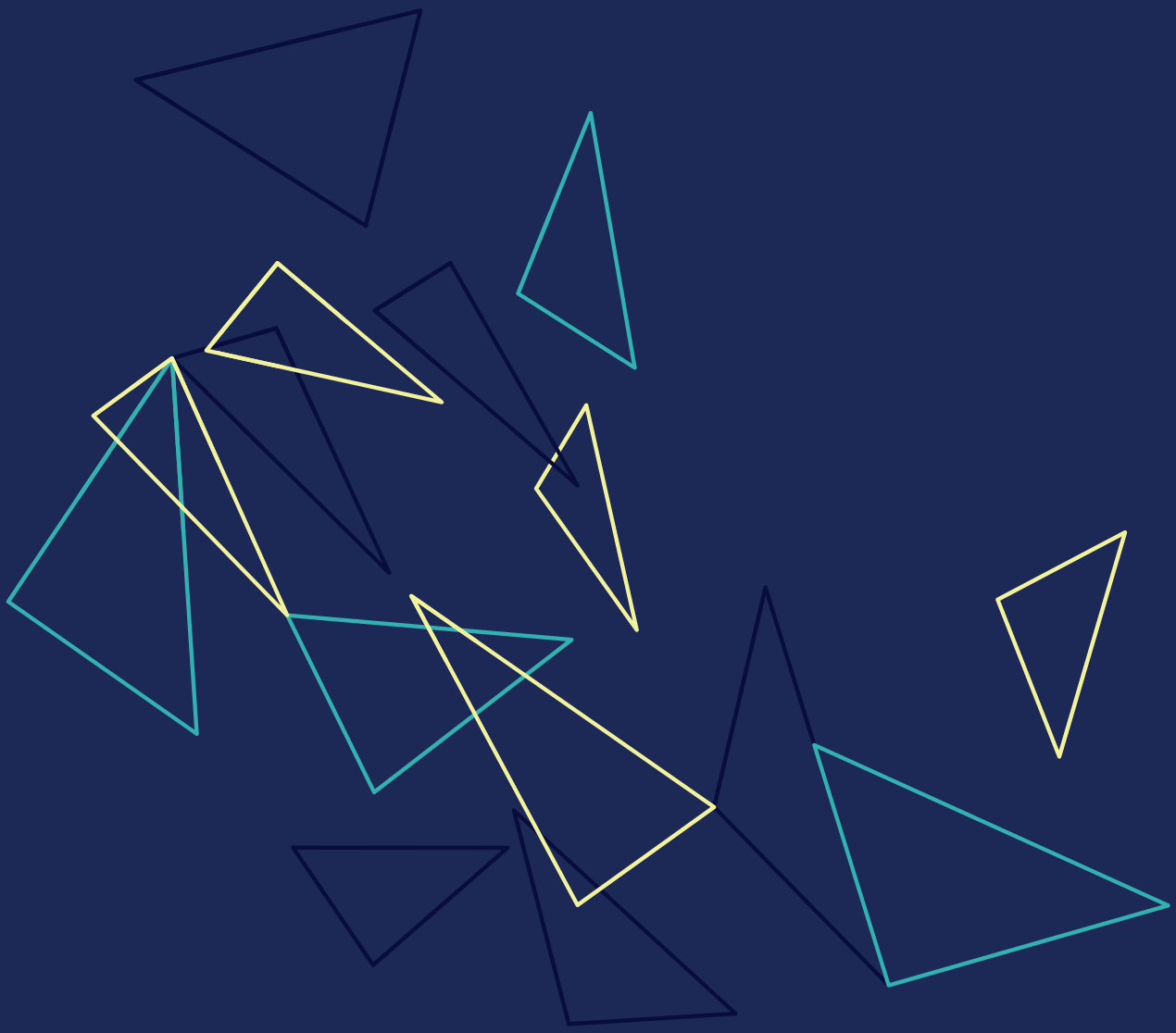
Platform Cooperativism

Platform Cooperativism developed in the past few years in response to the unilateral control of the sharing economy giants such as Uber have on the workers and AirBnB on the property owners on the platforms.

Platform Coop is a digital version of the time honored practices of cooperatives that enable the members to have democratic governance of the organization. Platform coops organize through digital platforms to provide the same services as large corporations and empower the members to be the owners of the platforms. Ride sharing coops such as Denver Green Taxi and Ride Austin have proven the coop can be competitive to the large and centralized alternative.

A kinder Digital Future

While the mainstream view of the future promoted by the digital giants seem to be bleak without alternative. The digital giant make the devices, control the platforms and use them to collect our data, analyze our behaviors, sort us into categories and influence our choices according to the buckets we are sorted into. The techo-determinism seem to make this future inevitable. However, communities around the world are taking actions collaboratively against this narrative of the future and bring a future of digital "good life" where the communities can exercise controls over the devices and the platform and individuals in the communities can explore the resources to take adventure and create indeginous innovations to keep on improve the quality of lives. That future is not just possible but desirable. We just have to work together to archive it.



What place for frugal technology in a permanently disrupted world?

Bogomil KOHLBRENNER, *Global Studies Institute (GSI) - University of Geneva*

Do simple and frugal technologies form part of the field of disruptive technologies? To what extent do we consider them to be innovations when we are instinctively more fascinated by technological advances at the forefront of their respective fields, from nano to the conquest of space, from quantum to genetics? In short, are frugal technologies, which we also refer to as "low-tech", disruptive and what is their place in a world attracted by interlocking complexity?

Despite a media focus on Artificial Intelligence, it would be wrong not to address simpler, more frugal technologies, whose impact on human society is, in fact, much broader. Quantum computers will certainly go down in history as a revolutionary achievement, but it is more modest changes due to their accessibility that have had an impact on the largest proportion of humanity on a daily basis. In this sense, they may certainly be more discreet but are just as disruptive.

Advanced technologies not only attract all the attention, they develop elective affinities with authorities and the elites. Always eager for funding, these innovations know how to make themselves desirable, giving those who control them the feeling of having a head-start on the power games of the future. Meanwhile, low-tech innovations do not attract as much attention because simplicity does not create such a buzz.

What low-tech has to offer, many in positions of privilege already take for granted. And above all, it may be thought that this only concerns, and is limited to, disadvantaged strata of society, holding the least prominent positions. That said, in this chapter we will not discuss cases where asymmetric parties to conflicts have access to technologies, the access costs to which have dropped significantly and who use technology against power (drones, computer viruses, biohacking, etc.).

Furthermore, the distinction between high-tech and low-tech is not obvious. By definition, any human production – material or knowledge – can be considered as "technology". As a result, frugal innovation, which consists in using the minimum of resources to efficiently meet a need, has always existed. From the outset, frugal innovation ignores anything superfluous to reduce the costs and resources needed to implement, use, maintain or even reprocess it as waste. It offers little or nothing by means of profit margins. By reducing complexity and the number of intermediaries, the number of anchor points where a profit could be made is also reduced.

Recent development of frugal technologies

Frugal technologies as a specific field are making a cyclical return to development, corporate social responsibility, and research and development circles. In the 1950s, questions were already being raised about "appropriate technology", a concern that became more pressing as technological revolutions accelerated.

Indeed, a parallel development in the consciences and discourses of the stakeholders involved can be observed. While the gap appears to be widening between these two types of innovation, which are located at opposite ends of the spectrum, it seems that we cannot conceive of one without thinking of the other. Frugal technologies are inseparable from high-tech technologies, in whose footsteps they tread. As we saw earlier in David Li's article, this reflects the logical curve of the dispersal of knowledge and investment, presenting an example other than mass prosperity.

The principles of frugality remain an inspiration for designers, as we have seen with the enthusiasm for "Jugaad" in recent years. The more we progress along Rogers' Innovation Adoption Curve, the closer we get to the principles of frugality: do more with less, and make it affordable for greater dispersion.

By seeking simplicity in design, use and maintenance, these principles also provide a response to problems of excessive consumption and waste of resources and life. Although frugality may not be the panacea, it represents part of the answer to the problems of wear and tear through ignorance and the disproportionality of our current systems.

In addition, frugality is particularly suitable for the living conditions of the poorest and most remote populations. On the one hand, it promotes empathy, but it is also a business strategy, recognising that advanced technologies cannot be built without a basic infrastructure and prior technological basics. This makes it possible for companies embarking on this path to develop new markets while maintaining their responsible image. Take Coca Cola, for example, whose market extends to the most remote corners of the world, and which – in its own way – offers an alternative in places where drinking water is not always accessible.

The development of small-scale solar energy – powering a light bulb or charging a mobile phone, or even a refrigerator and television in households with more resources – is an obvious example. This can be seen in regions where the authorities have not provided their populations with an energy infrastructure. Political promises and major development projects are beginning to take an interest in these operations, which are carried out on more reasonable scales.

If there is one area where the impact of low-tech is most obvious, it is energy. Solar energy is one example among many of how small-scale projects can be integrated into larger ones, the consequences of which can be potentially disruptive for society. Discreetly but surely, the integration of micro-solar and solar energy at the community level is being adopted in all rural regions of the world, bringing with it an essential infrastructure revolution for the development of more advanced technological breakthroughs. The adoption of these technologies is driven by users themselves, who want them and associate them with potential social progress.

Micro-solar, micro-hydro and wind energy is on the rise. These technologies have the advantage of being as close as possible to users, of reducing circuits, of providing something that did not previously exist, and of being simple enough to allow as many people as possible to understand how they work.

Despite the difficulties of integration into existing networks and intermittent energy, micro energy networks are being deployed. This approach makes it possible, for example, to solve intermittent problems by coupling renewable energies to water batteries by pumped storage or by creating water and salt batteries. These trends only partially address current and future energy issues. However, they share a potential for democratisation and can be used as a basis for further progress and complexity. As Gollwitzer also points out, these innovations tend to reduce the overall ecological impact.

Benefits that concern us all

Of all technologies, those relating to communication have had perhaps the greatest impact on development: transport, trade, and now perpetual digital connection. The latter, which has been made simpler, more robust and affordable for end users, is spreading in an unprecedented way.

These modest individual successes demonstrate that the characteristics of the frugal must be integrated with advanced technologies to ensure environmental sustainability, infrastructure sustainability and accessibility for the greatest number of people. The benefit is twofold: through reducing the impact of overconsumption of resources and through adoption by the greatest number of people. The new industrial revolution, that of a democratised and omnipresent cyber-physical system, however, invites reflection.

The 4th Industrial Revolution

To achieve this democratisation, this incorporation into our daily lives, and particularly in order to perpetuate itself, should the fourth revolution be accompanied by a fundamental paradigm? Is Mark Zuckerberg's disruptive maxim "move fast and break things" ultimately tolerable for this new iteration of the industrial revolution? By the way, in 2014, the giant's motto changed to "move fast with stable infra[structure]", thereby recognising the importance of a solid base rather than costly retrospective repair of the damage. The problems and difficulties of the three previous industrial revolutions are still deeply rooted in our time. We have made incredible progress in terms of population and health. An ever-increasing proportion of the population is benefiting from industrialisation and digitisation. But at what environmental cost? The sixth mass extinction is under way, to mention only one tiny problem among many that could be listed as a result of the growing activity of humans on earth. Thus, the recent improvement in humanity's quality of life, undeniable though it may be, is likely to become insignificant in the face of the systemic changes generated along the way. Although we can't go back in time – and would we really want to? – it is imperative not only to reduce the impact of our technologies, but also to work to revitalise our ecosystem.

Ambivalence and caring

Progress marches on relentlessly, however. It goes without saying that there will be benefits, just as there will be, with new uncertainties, legislative repercus-

sions, and collective crises of confidence. If the 4th Industrial Revolution is coupled with a paradigm shift, inspired among other things by a frugal approach, it may be possible to adopt caring technologies. Investors should even be able to demand it.
And caring is sorely needed. Indeed, technologies are, by nature, amplifying. Whether they are deployed on a large scale, automated, and used as tools, technologies are ambivalent. Take the knife, for example, which is both useful and deadly. We would certainly like knives to be neutral, but that would be to ignore reality. We must be aware that technologies, no matter how well-intentioned, irremediably carry contradictory constraints and potentials within them. We need to integrate an awareness of the systemic impact of technological, societal and individual choices and actions. In this sense, the principles of frugality make it possible, for example, to better integrate and solve environmental constraints.

Quality of life and capability

Addressing the problems of finite resources will not be achieved by limiting the poorest populations' (the very populations who are often experiencing population growth) access to developed countries' living standards. Imposing a lower quality of life where it is currently high also seems unrealistic. If collective choices prevent people from turning to these solutions, societies will be encouraged to turn to the principles of frugality. In addition to its ecological dimension, low-tech improves the well-being of stakeholders by establishing them as agents of the creation, repair and use of technologies.

Let us go beyond the idea of a rational, selfish human beings, purely focused on maximising their individual benefits. Nobel Prize winning economist, Richard H. Thaler, with his nudge theory, attacks this precise incomplete conception of homo economicus. We are then left with homo sapiens which often makes us overlook his penchant, homo faber. Our human condition sets us apart from the animal world by the fact that we are beings who flourish by making and creating things. By developing our expertise, by restoring our control, we give meaning to our existence, an object to our freedom and, in doing so, we collectively contribute to the construction of the society in which we live.

Developing appropriate and affordable technologies – in the financial and physical sense – is empowering for a large part of the population. Global networks should lead to the formative mission being launched by and for all. I am not referring here to the adage that you should teach a man to fish rather than give him a fish – the focus would be too much on the act of the giver. On the contrary, it is necessary to focus on the beneficiary and make them an actor of change. No tool is suitable unless it is fully appropriated by the actors concerned. Hence we need to ask whether the person even eats fish, and how the tools will be integrated into their socio-material environment. Appropriate technology is not only appropriate in economic and ecological terms. Frugal must be appropriate and appropriable. It provides opportunities for growth and makes it possible to be an actor of the tool and its environment.

People in need would not be the only ones to benefit from this. The empowerment that frugal technologies offer can be a source of satisfaction for even the most privileged, who, when they have everything, can ask themselves what they really need. The popularity of open-source and maker movements attest to this inherent need for humans to be actors in control of their

tools. It thus combines the improvement of quality of life through technology, well-being through individual development and, finally, collective contribution to the building of a better global society.

The ethics of care

Humans are born to live in communities, they aspire to belong to a group and contribute to the collective well-being. These values are reflected in the question of the ethics of care. "Care" is not limited to its medical meaning. It includes caring for others and caring for the environment. By promoting this ethics of care, we steer ourselves towards a morality that supports our deeply relational nature, rooted in our need and the interdependency of our lives between humans, nature and beyond. The moral ideal of equity between humans would be replaced by something more holistic, including considerations for the ecosystem and the cosmos.

More concretely, this would mean that the qualification of economic growth would no longer be based on predatory ignorance of the "hidden" costs and externalities that are left to others to manage. The principle of frugality, therefore, adopts the objective of development of technologies that take into account long-term viability not only for humans, but also for all forms of life. Usually, when we talk about disruptive technologies, about innovation based on destructive creation, we really mean a paradigm of breaking from temporary harmony to necessarily improve progress. From now on, we want to advocate a nourishing paradigm, the one that supports the substratum of life.

"Nothing is lost, nothing is created: everything is transformed".

Antoine Laurent Lavoisier

It might seem abstract, yet these are the very dynamics of our world which has developed towards such great complexity. Our technologies should tend towards a similar natural cycle: to ignore excess and superfluity, to be aware and adopt a global vision, to integrate complexity and simplicity.

What place for the frugal?

When we focus on meeting the basic needs of a large part of humanity, we have to ask ourselves how we design and develop all technologies. The quid pro quo is that this reflection concerns us all. The place for the frugal is here and now. Not necessarily in the sense of a decrease in growth, but rather in the sense of effective inclusion in the face of the complexity of technologies that seem to be increasingly out of our control. Without being able to provide a universal solution, it is necessary to be aware of the implications and intertwining of the complexity of technological fields. Frugality invites us to question what is necessary and to strive for simplicity. We hope that such an approach, guided by the principles of appropriate and caring technologies, will ensure development which is less destructive and provokes less anxiety. Frugality is also a guarantee that small changes, through the effect of scale, can lead to larger ones!

Examples of frugal technologies

These technologies address many basic issues, such as energy, health, water, transport and construction.

From a health point of view, simple solutions such as zinc-based rehydration solutions and jet injectors that allow needle-free injection have already saved hundreds of thousands of lives.

Innovative medical solutions are also designed using advanced technologies, but are frugal in their approaches. At one end of the spectrum of complexity there are micro-labs for the detection of malaria, which are economical and the size of a credit card (McBirney, Chen et al. 2018). These make rapid diagnoses easy, by simplifying the process. At the other end of the spectrum we find the Globaldiagnostix digital X-ray equipment developed by a consortium of Swiss universities and EPFL, which also aims at simplicity and robustness (Essentialmed 2015). Despite the fact that this is a high-tech device, it is designed to withstand the most variable conditions regarding energy, the environment, reparability and skills, striving for technical simplicity and ease of use, while reducing the overall cost through its components. It thus demonstrates that frugality of design and conventional technology are not in opposition to high-tech but are complementary to it.

In terms of water, silver or simple terracotta filters provide access to drinking water for a very large part of humanity. *Hippo Rollers* reduce water transport problems in rural areas.

With urbanisation, old transportation technologies are being given a new lease of life by combining with new ones to provide simple and efficient solutions. Examples of such technologies include electric bicycles (McCue 2018, Reid 2019) and electric scooters (Grouse 2018). Both simplify personal transport by reducing the complexity and framework of the vehicle, while still offering partial propulsion.

For more information :

Akram, S. (2019, 04.02.2019) "Africa Embraces an $8 Billion Solar Market for Going Off-Grid". Retrieved 15.02.2019, 2019, 2019, from https://www.ozy.com/fast-forward/africa-embraces-an-8-billion-solar-market-for-going-off-grid/92303.

Grouse, O. (2018, 02.11.2018) "Micro mobility barometer: the electric scooter boom." Retrieved 18.01.2019, 2019, 2019, from https://nexxdrive.fr/barometre-de-la-micro-mobilite-le-boom-des-trottinettes-electriques/.

Hipporoller (2018). "Hippo Roller." Retrieved 21.02.2019, from https://www.hipporoller.org/.

McBirney, S. E., *et al.* "Rapid Diagnostic for Point-of-Care Malaria Screening". ACS sensors 3(7): 1264-1270.

McCue, T. (2018). Global Electric Bike Market Is Still Moving Fast - E-Bike Offers Glimpse Sensors. Forbes. Reid, C. (2019). When Will E-Bike Sales Overtake Sales Of Bicycles? For The Netherlands, That's Now. Forbes, Forbes, Forbes.

Yee, A. (2018). Solar Mini-Grids Give Nigeria a Power Boost. The New York Times.

Reconciling innovative technology, growth and sustainable development

Interview with **Bertrand Piccard**, doctor-psychiatrist, lecturer and aeronaut. The man behind the first non-stop, round-the-world balloon flight, initiator and pilot of *Solar Impulse*, President of the *Winds of Hope* Humanitarian Foundation, United Nations' Goodwill Ambassador, and a pioneer of free flight in Europe. Interview by Bogomil Kohlbrenner.

How can we understand the tension at the very heart of technology, that it is both a source of problems and a solution to the problems that it has exacerbated? Sustainable development requires a global effort, collaboration between scientists, innovators, new technology enthusiasts, entrepreneurs, investors, public institutions, those involved in governance, not to mention users.

The aim here is not to promote the merits of protecting the planet, but to demonstrate that clean technological innovations can be used both to improve quality-of-life and to make profitable investments. To make the ecological ideals supported by the economic argument tangible. This is an attractive argument that is difficult for anyone to deny.

We had the chance to talk to Professor Bertrand Piccard, a Swiss scientist and modern-day adventurer, who succeeded, among other exploits, in travelling the world in a hot air balloon ten years ago, then again in 2016 on board the *Solar Impulse*, a fuel-free aircraft.

"If I came back from my expeditions and said that I was simply driven by a quest to set records, everyone would be on board. But that's not me. I prefer to abandon the language of the adventurer to convey the messages that are close to my heart."
Bertrand Piccard

Definitions

What is sustainable technology? This vast field includes all technological solutions to protect the environment. Previously triggering great resistance, such technologies were rarely implemented because of their cost. In the past, solar energy was too expensive, as was electric mobility and home insulation. Finally, there was a huge gap between industry and ecology. The guarantors of growth tried to show that solutions made it possible to reconcile ecology and growth, but with the same financial barriers, or even a decrease in the quality of everyday life.

Now, most of these technologies have not only become affordable, most are even profitable! The price of solar energy has been slashed by ten. So has the price of wind energy. All new technologies regarding mobility, home insulation, heating, air conditioning, lighting, smart-grid, distribution, etc., have, fortunately, become profitable.

Examples from the Global Clean Technology Alliance

Bertrand Piccard demonstrates this profitability with some of the latest technological discoveries certified by the Global Alliance for Clean Technologies. Founded when he returned from his Solar Impulse experience, when he saw the enormous potential of renewable energies, this organisation promotes clean, efficient and, above all, practically achievable solutions.
These are just some examples: an industrial process that produces stainless steel using 99% less water, while being 91% cheaper. Ceramic blocks to store heat and transport it from where it is lost to where it is needed. A solar energy system to desalinate seawater, with variable supply currents. An air conditioning system using seawater, which takes water at a depth of 900 metres at 5°C to cool the air on the surface of large hotels or hospitals, as has already been done in French Polynesia, before returning the water to the ocean without destabilising its temperature. Not only is a lot of energy saved for air conditioning, but real financial savings are also secured.

As a final example of the immediate usefulness of these innovations, a member of the organisation submitted a solution that has been certified for combustion cars. Well aware that, despite the introduction of electric cars, the world's fleet of cars will not change immediately and combustion engines will be on the road for at least another ten to fifteen years, the proposed solution is a simple box to be mounted against the engine. This box electrolyses the liquid, putting some hydrogen into the engine combustion chamber, which improves combustion while reducing particulate emissions by 80% and fuel consumption by 20%. This is clearly a breakthrough technology, making huge improvements to what exists without having to switch to electric cars right away.

The paradox between "development" and resource "reduction"

At this stage, let us couple the definition of technologies with the notion of "development". We cannot continue to have growth that uses more and more natural resources in order to have more and more objects to throw away. This is quantitative growth. This growth is not possible because our world and its resources are not infinite. If we adopt new technologies that consume more energy and natural resources than before, we will be on the road to ruin.
Economic development is viable when it is based on

qualitative growth. Quality can be improved by replacing things that pollute, that are inefficient, that waste wastes energy, that waste water, that waste natural resources, with systems that are efficient, that save energy, and that save raw materials. And that, Bertrand Piccard makes no bones about it, it is the market of the century!

Increasing quality, not quantity, means, for example, introducing smart-grid infrastructures that allow cities to become carbon neutral rather than wasting all the energy they consume. There are many solutions to transform and improve what we have, without having more, but having better.

In financial terms, this qualitative growth represents both profit and jobs. Bertrand Piccard is currently working on this with the Solar Impulse Foundation. Its objective is to identify one thousand of these solutions, which are economically, financially and industrially profitable to protect the environment.

Update on the situation in 2018

More generally, Bertrand Piccard's enthusiasm is counterbalanced by an extremely contrasting perception of the current situation with regard to technological progress. When you talk to people in Europe about what is happening in California, they don't believe it. In California, they claim not to be 20 but only two or three years away from autonomous cars. California has decreed that by 2045, 100% of electricity will be renewable. In other places, they do not even dare to give 2045 as the date to remove diesel engines from cities.

Another example, the King of Morocco, Mohamed VI, declared that by 2030, 52% of Moroccan energy will be of renewable origin. And they will reach that goal, because they have already started. Meanwhile, in our countries, with governments that change regularly, the goals are set at 2050, but we know full well that we will not reach them. Politicians are lauded until they lose the next election, and their successors will not achieve these objectives, because they won't be there in 2050 either. And when we do get to 2050, we will see that nothing has been done.

"The adventure of the 21st century is to use human creativity and pioneering spirit to develop the quality of life to which current and future generations are entitled."
Bertrand Piccard

Towards a re-industrialised Europe

Should we despair of what is happening in Europe? Certainly not, there is significant potential. In Europe, there are many patent applications, European research is fantastic, the European Commission is very involved. However, once patents are filed, inventions do not reach the market, they are not used. That is, start-ups come up with interesting solutions, but these are not recognised. It is crucial to ask what it would take to make this happen. Innovation must be driven and not just pushed. We push it with grants, subsidies, incentives, pitches, lots of stuff ... but that's not enough. A need must be created to bring these innovations to market. And the need can only be created by regulation. This is the legal framework.

In practical terms, if so much CO2 is allowed to be put into the atmosphere without constraint, new catalysts that can be developed to capture CO2 as a raw material instead of releasing it into the atmosphere, will simply not be developed, because this is not going to be profitable. No one will be interested. Once there is a CO2 tax, it starts to have a value. And hence, CO2 technologies will develop. If there is a CO2 tax, we will have to be more efficient in terms of heating, air conditioning, engines, etc. Today, industrial electric engines consume 30% of the world's electricity. We can have electric engines that are 60% more efficient, identical to those used on the Solar Impulse, but without a CO2 tax, no one will replace the old engines.

So, if we really want to re-industrialise Europe, if we really want companies to come up with breakthrough inventions and to bring these inventions to the market, we need a regulatory framework that includes very ambitious energy and environmental policy objectives. Otherwise, technologies will not emerge.
In this sense, according to Bertrand Piccard, the Swedish example is very interesting, in that the country introduced a carbon tax of €35 per tonne. At first, industry screamed: "this is the end, we will no longer be competitive!" But today, thanks to this carbon tax, Swedish industry is more competitive than ever in exporting because it has been forced to use new technologies.

Each to their own contribution.

During the interview, Bertrand Piccard said he was convinced that those who are the flag-bearers of change are those who have cost-effective solutions for protecting the environment. From this point, politicians must take over, and impose standards – because, let's face it, it is the legal framework that will change the situation. They then need to provide funding, so that industry understands that it is absolutely profitable to invest in sustainable technologies.
Let us look in more detail at the awareness message that each group of actors should be expounding.

Call for investors

Since the main argument is based on profitability, what about the financial world in particular? Fossil fuels are an exhaustible resource. Imagine: you have investments in fossil fuels, and you know that an oil company's share is worth 100 (to simplify the calculation) because there is a number of reserves that is worth 100. What happens after the introduction of a carbon tax? Solar and wind energy will become much cheaper than oil. Electric mobility will take place. There will be energy efficiency in buildings. Yes, but it is not simply a question of costs. At the same time, you learn

that these oil reserves in which you have invested will never be fully used.

Therefore, if the company was valued at 100 because it has reserves at 100, but only 30% of these reserves will be used, it means that the shares will collapse from 100 to 30. And when the population realises this, and if they realise it quickly, the result will be a stock market crisis. We will face the biggest financial crisis in the history of mankind. Today, oil companies have junk assets, much like the sub-primes in 2008.

How can this be avoided? Oil companies need to diversify. For the sake of their survival. They must invest in energy efficiency and in all renewable energies. If oil companies have time to retrain, there will be no financial crisis. Even to save the current financial system, it is essential to encourage these fossil energy companies to adopt the logic of renewable energies. New technologies will not only be alternative sources of energy, but also valuable assets.

In reality, if you say that to oil investors they'll laugh, saying that oil has represented 80% of the energy mix for 50 or 60 years, and that it will continue to do so no matter what. The financial story is, however, different. It will not continue. Suddenly, their shares will be worth 30% of what they are worth today. Even if there is a need for oil. It is a stock market problem much more than an environmental problem.

Call to entrepreneurs

To take an analogy, horses clearly still exists, but this did not prevent those who had horse-drawn carriage companies in the cities from realising that society had switched to using cars. There will always be oil. There will always be a niche that will use oil, but society will have moved on.

It is a trend, a trend that will never be stopped. The choice remains to oppose it and be doomed to disappear, much as Kodak did, or to embrace change, to diversify, in the way that LG did. LG moved away from fossil fuels to invest in energy efficiency services for their customers. This is a new business model where you earn a lot more by selling less, because you're more efficient. And it is this value that pays off.

Another anecdotal example can be gleaned from a discussion between the adventurer and an Italian industrialist at the head of a heavy machinery sector, who voluntarily invested his own capital to reduce his CO2 emissions. Results: 18% profitability per year, 18%! Environmental protection pays well. That is what we need to show people. It is as simple as that.

Call to citizens

On their own level, every individual can make this kind of investment. Bertrand Piccard himself has insulated the roof of his house and installed a heat pump, thus dividing his annual expenses by three. If he had left the money in the bank, he told us, it would be paid at zero, if not at a negative rate. It seems obvious to him today that equity capital must be used to be more efficient. Regardless of people's values and convictions, we can do better for the environment (and our wallets) by looking at everything that is profitable in our daily lives: LEDs, electric cars, house insulation, heat pumps ...

Bertrand Piccard also advocates for the establishment of a regulatory framework that would restore balance. You cannot stop people from travelling by plane, but

you can introduce a tax on kerosene. It's insignificant in terms of price of the ticket, but it can restore the balance. If someone goes shopping in Barcelona rather than staying in Geneva, they can easily afford a 3 or 4 Swiss francs tax on kerosene tax on their EasyJet ticket. Until this takes place, CO2 is added to the atmosphere without adding the means to remove it. A tax allows these externalities to be deducted. Were there a carbon tax, it could be used as way of financing the reduction in CO2 in the atmosphere, to develop solar power plants instead of coal-fired power plants. And until that is done, we will see that technology can destroy humanity, just as it can save it. Ultimately, it is what we do with technology that counts. The same technology can destroy a city or produce electricity. The same technology can be used to monitor people with artificial intelligence or find missing people. It can simultaneously be a factor of social development or a factor of dictatorial surveillance. It depends on what you do with it.

Moreover, what we tend to do is to always try to outsource costs in such a way as to pass them on to someone else, rather than to accept them and make them a factor to stimulates efficiency. Today, the externalities of nuclear energy and the oil industry are having a clear unbalancing effect on the energy market. In practice, fossil energy is subsidised while renewable energy contains all its costs within itself. It is said that nuclear energy is cheap, without talking about the state's insurance in the event of a disaster – and we have seen how much it has cost Japan and Ukraine – or CO2 externalities. We want to let the laws of the market decide; those same laws of the market that are totally biased by externalities that are not taken into account. The world today is so badly managed that if a company were managed in this way, Bertrand Piccard's CEO would be in prison.

Moreover, given the ecological, social and economic limits of our system, not to mention the fear of any change, the population cannot rely solely on its leaders. We have to acknowledge the inertia of the system and the loss of confidence. The result is significant abstention from the system. Voters believe that there is no point in voting between plague and cholera, that it is futile to get involved.

Call to decision-makers

This remarkable abstention from the electoral system is accompanied by an increase in populism. Why? Populism arises from a reaction to the lack of consideration of the leaders for the population: no explanation, no vision, no goal or resources are given. And we say to people, "Vote for us, we are better than the others anyway". And when a populist party arrives, this view means that the population will vote for them. Today, there is an international shift in favour of populism in response to the lack of vision and incompetence of the leaders of recent decades.

World leaders and large corporations no longer inspire confidence. The political world needs to get its act together. One of the major challenges ahead is to successfully restore trust. This is certainly wishful thinking, and Bertrand Piccard is unfortunately aware of this and almost apologises for himself for dwelling on more concrete things than politics. However, his efforts are focussed on finding solutions that can both protect the environment and allow business to flourish, as many people are interested in a positive financial recovery.

Based on his own experience, Bertrand Piccard can

only advise political leaders to show vision, and explain why and how this should be achieved. According to him, if political, economic or industrial leaders dare to take the step, they will be followed and will become heroes.

Unfortunately, these leaders are currently considered by the population as parasites who only think in terms of short-term profit. They have no respect for others. Benevolence seems to be absent. However, when there is benevolence for employees, consumers, customers, the environment, etc., business becomes better by finding its social utility. A group in France was recently created to reflect this cause of benevolence. Because it inspires confidence and attracts interest that would not otherwise have been crystallised. We must advocate a return to basic human values – not to mention spirituality – of benevolence, respect and humanism. It is something that has been completely underestimated in our world for decades, but is now beginning to make a comeback, almost in the sense of a business argument! Bertrand Piccard admits, if this argument could also become a political argument, it would be magnificent, although we are still far from this point.

A call to the generations of today and tomorrow

In this regard, would education make it possible to move in this direction, by communicating our knowledge and values to future generations? Successful education is a great thing, except that most people are currently being taught the idea of a consumer society, a polluting society, linear rather than circular consumption, and wasted energy. It is a sign of wealth if you have air conditioning. Education in this sense will not change the world. Can we afford to wait for the new generation, when climate change will mean that in 25 years we will already have temperatures which are 3°C or 4°C higher? We must act today! Regardless of education, we must start now. Not only for the good of future generations, but also for the benefit of our present generation.

Paradigm shift

Born into a family driven by humanism and aware at a very young age of the need to preserve the planet, Bertrand Piccard saw his father carry the same message for the protection of the environment. However, at the time, there was no cost-effective solution. People were not ready to listen, and even less inclined to apply it. Faced with this situation, Bertrand Piccard is convinced that the paradigm must change by making environmental protection profitable.

"Pioneering spirit is not about finding new ideas, but about getting rid of certainties and habits that keep us trapped in old ways of thinking and acting."
Bertrand Piccard

In 2002, when he started working on Solar Impulse, and published the first articles on the profitability of environmental protection, he was laughed at: "You're starting to create environmental capitalism! The Greens are not about making money, they are about saving the planet!" But he believes that we will not save the planet if it costs us money. We'll only save the planet if it makes us money.

And here it is, the great paradigm shift: environmental protection is the industrial market of the century. Whether we believe in the human origin of climate change or not, whether we respect the environment or do not, whether we have empathy for humanity or not, ultimately this does not change the fact that these technologies will protect the environment, and that they are profitable.

You have to speak the language of those you want to convince. Sometimes you have to go back to basics. When we talk about "capitalism", and when we talk about "capital", this capital is not only financial. It is also human. And it is also environmental. Therefore, if we really want to be capitalist, in the true sense of the word, we must integrate human capital and environmental capital.

Conclusion

Bertrand Piccard is pragmatist and knows full well that the world cannot be changed with good intentions. It is not enough to say that nature is beautiful and that it must be protected. Business leaders think about the salaries to be paid at the end of the month; analysts think about quarterly reports. And it is for them that innovative solutions must be offered. Of course, ethical criteria, such as responsible and local consumption, are fundamental. But that's another debate. You can boycott illegally deforested wood from the Amazon in favour of certified wood. You can boycott illegal fishing. You can stop buying products from the other side of the world and try to consume more locally. But not everyone will do this. But when we talk about profitable technologies, everyone can understand that.

Adventure is a state of mind towards the unknown, a way of conceiving our existence as an experimental field in which we are obliged to develop our inner resources, to climb the path of personal evolution and to assimilate the ethical and moral values that we need as travelling companions."
Bertrand Piccard

What about those of us who really care about ecology? Make your attitude contagious, encourages Bertrand Piccard. If, in addition to profitability, some people are interested in protecting the environment for its own sake, so much the better. They are people who have respect, empathy, humanism, or humanity. This is perhaps the most important dimension, even if it is not the most widespread. As long as their attitude can motivate others to think as they do. For others, profitability may be the only motivation. Even those who deny climate change will be able to adopt sustainable technological solutions. Maybe we shouldn't even tell them that it protects the environment, to keep it simple, just that it's an obvious economic advantage.

How do we hold AI itself accountable? We can't.

Dr. Joanna BRYSON, *University of Bath, England, United Kingdom, European Union*

Artificial Intelligence (AI) is often presented to us as another race or gender of human that has growing superhuman capacities. It is natural therefore that many ask how we can integrate these new individuals into our society and our system of justice. Unfortunately, this presentation is entirely erroneous. Intelligence is an attribute of an agent, not an agent in itself, and artefacts with or without this attribute cannot be dissuaded by human justice. Hu- man justice is uniquely designed for maintaining societies of organisms like ourselves

This report is a short precise of a formal academic article on legal personhood for AI that I wrote with two leading law professors in legal personality, Tom Grant of Cambridge University, and Mihailis E. Diamantis of University of Iowa. Since they each had far more influence on the article than I did, I can sincerely and humbly say that that article is a great paper that I think everyone should for themselves. The title is Of, for, and by the people: the legal lacuna of synthetic persons, and it appeared open access (thanks to fees paid by the non-profit University of Bath to the for-profit publisher Springer) in the journal Artificial Intelligence and Law 25(3):273-291 in September of 2017.
I recently received an email about that paper, and I repeat the letters we exchanged here. First, the (anonymised) initial email: "I'm writing to you in view of your article 'Of, For and by the people: the legal lacuna of synthetic Persons'. What are some of the mitigation measures that should be in place to ensure synthetic persons are legally accountable for their acts in case they are granted electronic personhood?"

The reason I ask is that the point of our article is that there is no way to ensure that a synthetic person can be held legally accountable. It does not matter whether you mean a 'synthetic person' to refer to a robot, or to the legal fiction that is used to make a corporation appear like a person. The only way to ensure that law is stable is to have a human be accountable for the actions of an artefact, and that same human be the one in control of the artefact's behaviour.
In this report, for clarity, the term human will always refer to a biological entity of the species Homo sapiens. However a person will be a person recognised as such by the law. Some humans are not person, because they are not competent to operate in the context of the law (e.g. infants or those with severe dementia), or because they are not recognised by the law as persons (something that might happen for example to a member of an ethnic minority living under an autocratic regime). But some non-human entities are legal persons, such as companies, and sometimes religious idols, including in one case a river. For all of these non-human entities, legal personhood is attributed (assigned) to the entity because it is legally convenient, and there is a sense in which justice can be upheld. Idols are only assigned personhood in that they are moral patients, that is, that they need to be protected as if they were a human. There are two reasons this makes sense for an idol:

• Real humans have been shown to suffer grievous harm when the idols do. This is partly because the idols are of great religious significance and therefore are part of both individual and community identity.

The other part is the second problem:

• The idols are unique and irreplaceable. They are either ancient artefacts that require preservation, or as I mentioned in one case the artefact is a river, which can be said to be killed if the pollution in the river is so great that the life depending on the river is destroyed.

An AI system might be unique, but if so, that would be a design decision. All AI is by definition an attribute of an artefact, and if it is a digital artefact, any intelligence on it, for example its individual memories, can be backed up and stored. Whoever built the artefact could likewise choose to use mass produced, perfectly replicable components. So unlike humans, rivers, or ancient religious artefacts, if an artefact with AI is unique that was a decision taken by a contemporary

individual who could easily have made a different decision and protected the intelligent system they were building. What we recommend in our article is that all legal commercial products including AI should be manufactured not to be unique, if there is any concern that humans would suffer were they to lose access to the AI in that artefact.

It should be said first that not every legal system recognises idols (or even corporations) as legal persons, and second that I learned the above about idols from an excellent paper by Solaiman, which was also core to the arguments my colleagues and I made in our paper. Finally, it should also be said that the arguments I make below about why AI cannot be held accountable through this mechanism also apply with increasing frequency to corporations. "Shell companies" are those founded only too deceive the law and remove the threat of legal action from humans or companies that the humans in control really care about. I'm sure the janitors of a shell company goes bankrupt, but increasingly some actors are happy to (for example) build buildings with the sole purpose of having the project go bankrupt and thus serve for money laundering. They may also enjoy as a power move or benefit politically from removing whatever attribute of a city had previously been built on the location of the bankrupt building, but that's only tangentially relevant to the question at hand.

There are two necessary conditions for an entity to be a legal person.

1. First, that entity must be able to know about and be able to execute the law on their own behalf. This is why animals are not held to be legal persons, though note that we do routinely allow infirm humans (and in some countries, idols) to be represented by others.

2. Second, the penalties of law have to serve as dissuasion to the entity. This is where shell companies (as just described)—and AI—fall down.

Although many people think the purpose of the law is to compensate those who are wronged, what the law mostly does really is to maintain order by dissuading people from doing wrong in the first place by making it clear what the costs are for doing wrong. If they do do wrong, they are forced to pay those costs, with the hope that this more immediate experience of the dissuasion will stop them from doing it again, or sometimes they are just forever prevented from free action either by being jailed for life or executed. Of course, sometimes part of the dissuasion includes recompense to persons wronged, for example the return of property, money, or even the granting of money to compensate for injury or time.

Humans are incredibly social beings. One consequence of that is that our society and self-image has co-evolved with our sense of justice. So often people do feel compensated when they see someone else dissuaded. But having the murderer of your partner jailed or executed by no means brings your partner back to life. It is good for the victims that they can feel a sense of peace, and perhaps they really do gain greater security if it is publicly known that the last person who wronged them was penalised.
But essential to all of this is that the entity that committed the crime is dissuaded from doing so again. This is also why tort settlements against companies can be outrageously high. When an elderly woman was awarded an enormous settlement after recei-

ving third degree burns from McDonald's coffee, it was not because the woman needed the money, but rather because a smaller settlement would not have persuaded McDonald's to reduce the temperature at which it kept its coffee. Similarly, the penalties the EU proposes against tech giants who violate EU privacy or competitiveness laws are set not for redress as much as for dissuasion.

I mentioned that humans have co-evolved with our intuitions about justice. Think about it: why is it punishment to put someone in jail, or label them a felon, or take away their home, or to fine a person (including a corporation) for an enormous amount of money? It is because humans have an enormous systemic aversion to isolation and losing power. We share this with other social species—even a guppy will die of stress if it is isolated from its society. Again, just as with uniqueness, if AI were to also display this aversion, it is a consequence of design decisions taken. In fact, there are fantastic amounts of extant AI and none of it minds at all that it is entirely treated as a tool, subordinate to human will, turned off, traded back in to Apple for the new iPhone, etc. Humans have so much trouble understanding how an intelligent entity could not feel betrayed by such action that they refuse to recognise vastly superhuman intelligence as intelligence. Can you do arithmetic as well as your phone? or spell as well? Even if they do recognise it, then they make up a new term for intelligence that would mind, like 'conscious' or 'general'. Unfortunately, these terms already have other meanings entirely irrelevant though sometimes coincidental to the real matter at hand here.

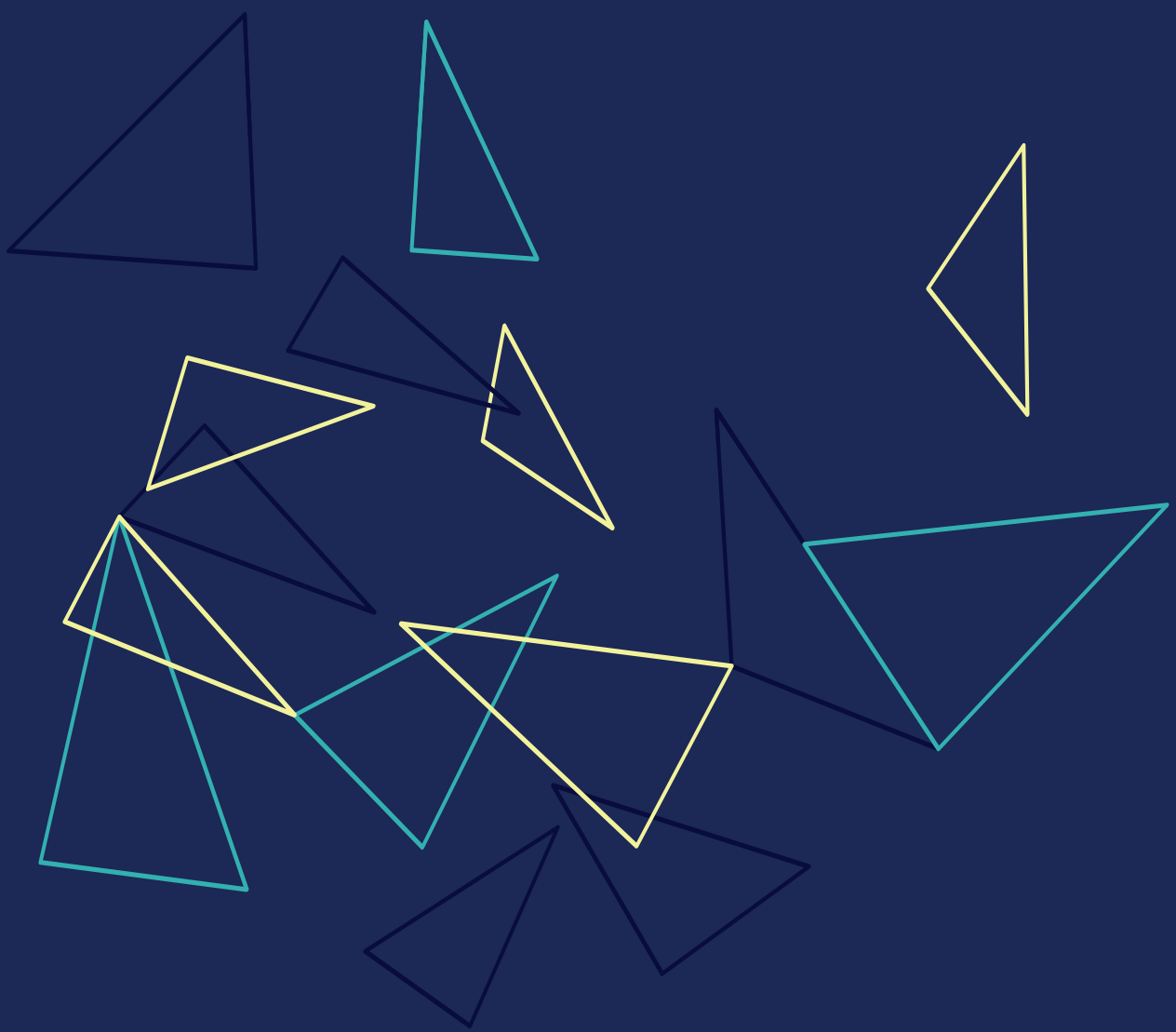
What matters is that none of the costs that courts can impose on persons will matter to an AI system in the way the matter to a human. While we can easily write a program that says "Don't put me in jail!" the fully systemic aversion to the loss of social status and years of one's short life that a human has cannot easily be programmed into a digital artefact. Even if we could program it, what right would we have to make something that will be bought and sold capable of suffering? But generally speaking, well-designed systems are modular, and systemic stress and aversion are therefore not something that they can experience. We could add a module to a robot that consists of a timer and a bomb, and the timer is initiated whenever the robot is alone, and the bomb goes off if the timer has been running for five minutes. This would be far more destructive to the robot than ten minutes of loneliness is to a human, but it would not necessarily be any kind of motivation for that robot. For example, again of a smart phone, if you added that module to your smart phone, what other components of that phone would know or care? The GPS navigator? The alarm clock? The address book? This just isn't the way we build artefacts to work.

Law has been invented to hold humans accountable, thus only humans can be held accountable with it. As I mentioned when I was describing shell companies, even the extension of legal personality to corporations only works to the extent that real humans who have real control over those corporations suffer if the corporation is to do wrong. Similarly, if you build an AI system and allow it to operate autonomously, it is essential that the person who chooses to allow the system to operate autonomously is the one who will go to jail, be fined, etc. if the AI system transgresses the law. There is no way to make the AI system itself accountable.

Having said that, it is quite easy to make the people (human or corporate) who use AI accountable, more so than within ordinary human organisations. What we can do is require that the way that any intelligent system is built— and if it has machine learning, is trained—is fully documented, and that that documentation is encrypted and secured. Further, many of the operations of the system—its decisions, and what it perceived when it made those decisions which determined those outcomes—can be recorded, a process that is called logging. This can make the system accountable in the sense that you can do accounting with the AI system, just like you can use books to make a company accountable for its finances. But the true executive of that company is the one that has to be held responsible with the evidence gathered from these methods, whether the conventional books of accounting, or the digital logs of AI.

In our article, for the sake of argument, we admit that some people might possibly find that there are rewarding aspects to building unique, suffering AI that really would benefit from legal personhood. But what we argue is that the probable costs of social harm from corporations and individuals evading their responsibilities by offloading them to AI far, far outweigh any benefit that would come to society by creation of such a vulnerable and needy form of AI.

I mean, think about it. Why would we want to motivate corporations to fully automate part of their business process (that is, get rid of any human employees) by allowing them to cap their legal and tax liabilities at the costs or establishing their new artificial legal personality? The European Parliament (EP) asked the European Commission (EC) to consider this possibility; fortunately it didn't take the EC long to consider and dismiss it. Probably part of the motivation of the EP was European Car Manufacturers lobbying because they are worried about competing with Apple and Google in the driverless market, because those tech giants have more money than they can legally spend, so are fully willing to take on all liability for their driverless cars. The injustice of this vast economic inequality does need to be addressed, but not by exposing European Union citizens to bazillions of new shell companies on wheels.



Contributors

Audrey AZOULAY
Director-General of UNESCO

Marylaure BLOCH
Global Studies Institute - Genève University

Joanna BRYSON
University of Bath, England, United Kingdom, European Union

Damien de CHILLAZ
Vice-Président B2B Platforms & New Business Models, Capgemini

Fabrice EPELBOIN
Teacher at Science Po Paris

Pierre GUEYDIER
Director of Research, OPTIC**Bogomil KOHLBRENNER**
Global Studies Institute - Genève University

Pr. Dominique LAMBERT
Professor of Philosophy of Science, University of Namur, Member of the Royal Academy of Belgium

David LI
co-founder and director of Hacked Matter, Maker Collider and Shenzhen Open Innovation Lab

Pr. Helen MARGETTS
Professor of Sociology and the Internet at the Oxford Internet Institute

Lionel MAUREL
co-founder of La Quadrature du Net

Edouard MORIO de L'ISLE
Consultant Blockchain & B2B Platform Consultant, Capgemini

Adrian PABST
Dean of Political Science faculty - Kent University

Pr. Bertrand PICCARD
Doctor-psychiatrist, lecturer and aeronaut

Romina REBOIS
Consultant

Claire SOMERVILLE
lecturer, international affairs executive director of the gender centre, Graduate Institute Geneva

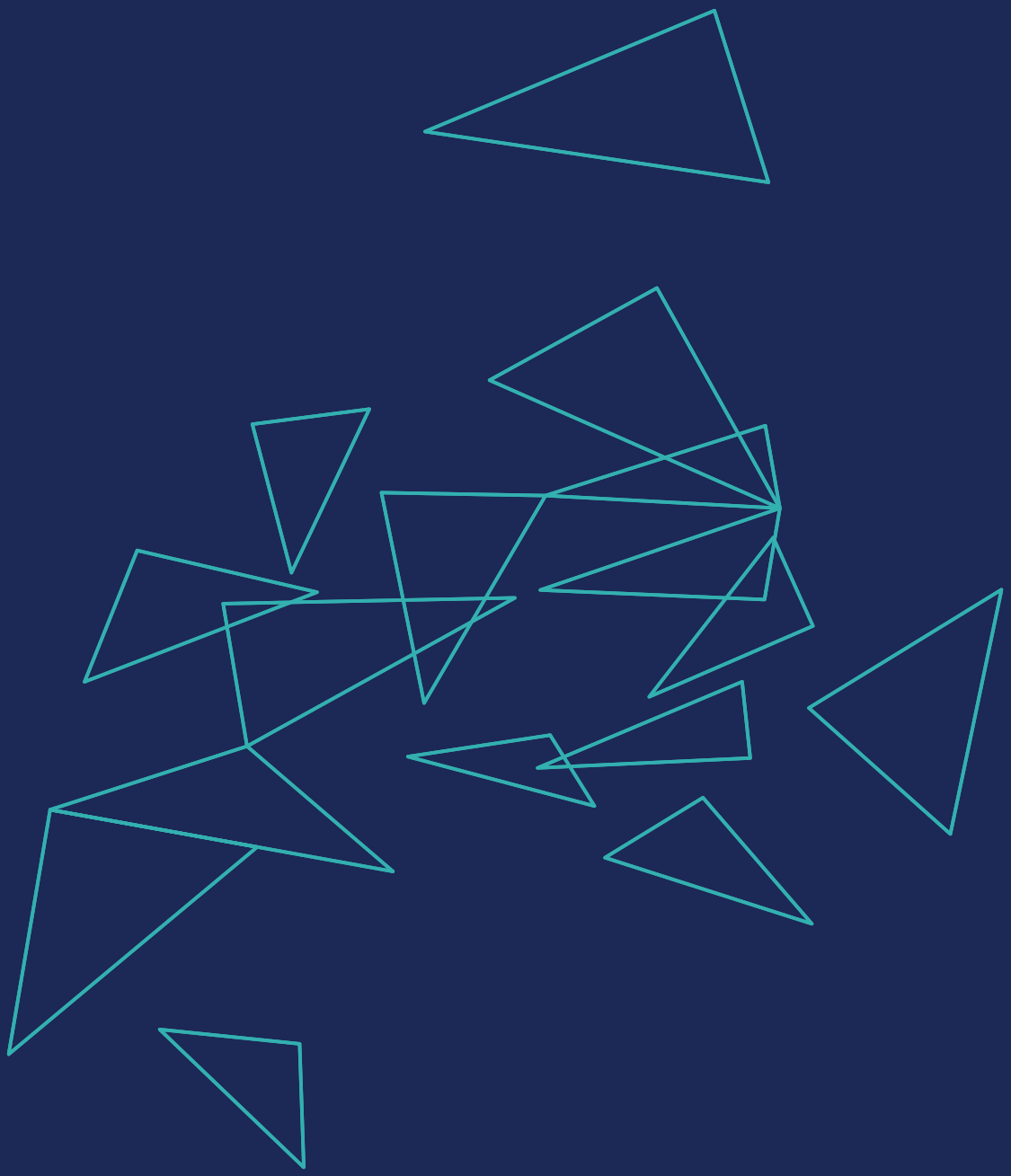
Laure TABOUY
PhD Neuroscience

Mariarosaria TADDEO
Researcher at the Oxford Internet Institute and Deputy Director of the Digital Ethics Lab, University of Oxford

Bernardas VERBICKAS op
Vilnius University, Law Facukty

Izadora XAVIER

Crédits
Cover : © Getty Image - Qi Yang
Audrey Azoulay / © UNESCO - Christelle Alix,
Helen Margetts / DR , Damien de Chillaz / DR,
Dominique Lambert / © Unamur Benjamin Brolet,
Mariarosaria Taddeo / DR, Bertrand Piccard / © Solar Impulse | Stéphane Gros
Solar Impulse / © Solar Impulse | Revillard | Rezo.ch
Others crédits photo : Shutterstock / DR



OPTIC