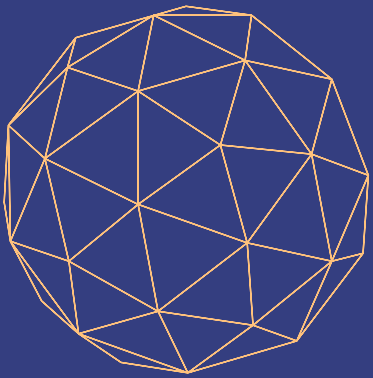
Issue 4 - April 2022

Good In Tech RESEARCH NEWS

Rethinking innovation and technology as drivers of a better world for and by humans



Governance & and responsible Artificial Intelligence

VALENTINE CROSSET, BENOÎT DUPOND

Governing extremist speech by algorithms

CHRISTINE BALAGUÉ, ZELING ZHONG

The Role of Consumer Perceptions of the Ethics of Machine Learning in the Appropriation of Artificial Intelligence-Based System

MELLET KEVIN, BEAUVISAGE THOMAS

Where does the value of data lie?

Institut Mines-Télécom Business School
TELECOM Paris TELECOM SudParis INSTITUT POLYTECHNIQUE DE PARIS
FONDATION DU RISQUE Louis Bachelier
SciencesPo ÉCOLE DU MANAGEMENT ET DE L'INNOVATION

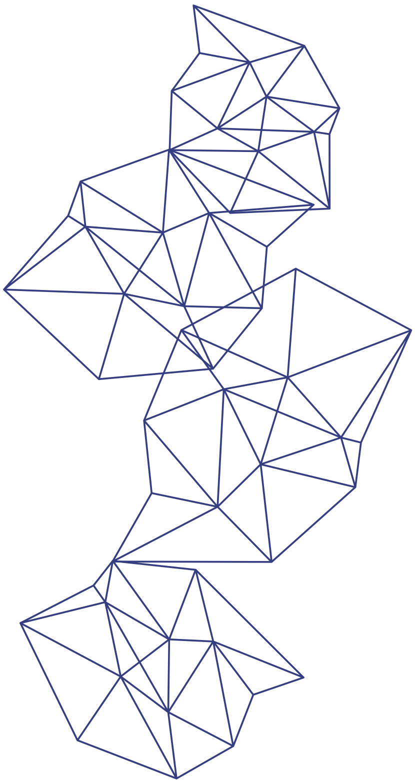
GOOD IN TECH VISION

Good In Tech main objectives are to create knowledge around four research areas and to contribute to the dissemination of this knowledge not only in academic and pedagogical spheres but also to corporations, decision-makers, regulators and the general public.

To this end, the Chair aims to create and develop an ecosystem of interactions between research, companies, students from the two partner academics and political institutions, civil society in order to raise awareness of all stakeholders on this new paradigm on responsible digital technologies and innovation.

The chair also aims to develop international partnerships, particularly in Europe, to share the issues of responsible digital innovation with international committees.

Finally, the Chair aims to share the results of academic works and debates it organizes with national and European political institutions in order to inform and influence public policies.



Governing extremist speech by algorithms

Valentine Crosset, Benoît Dupond

Valentine Crosset

Valentine Crosset holds a PhD in Criminology from the University of Montreal. She is currently a postdoctoral researcher at Medialab since November 2020, under the direction of Dominique Cardon. Her research topics, funded by the Good in Tech chair, focus on content moderation and sociology of the digital.

Benoît Dupont

Benoît Dupont has held the Canada Research Chair in Cyber Security since 2016. From 2006 to 2016, he held the Canada Research Chair in Security and Technology. He is a full professor at the School of Criminology of the University of Montreal and Scientific Director of the Integrated Network on Cyber Security (SERENE-RISC), which he founded in 2014. He also sits as an observer representing the research community on the Board of Directors of the Canadian Cyber Threat Exchange (CCTX).

WHY IS THIS TOPIC IMPORTANT ?

The aim of the article is to provide a critique of the concept of algorithmic governance by starting from a **concrete case**, that of the **moderation of violent extremist content**. Indeed, this research has two focuses: algorithmic governance on the one hand and new techniques for managing the control of the terrorist threat on the other.

This article is part of the whole debate on algorithmic governance. The aim is to develop a critical approach to this governance because if we look at how the topic is discussed in the literature, algorithmic governance is generally considered to be extremely powerful, capable of governing and sorting our lives. This approach implies the idea that algorithmic governance leads to new opaque regimes of control, population management, discrimination, etc.

According to the authors, although these arguments are very powerful and interesting, **algorithmic governance techniques remain rooted in traditional rationalities of surveillance and population management**. The idea is therefore to assert that algorithmic governance is not just limited to its purely technical form.

It is therefore a question of opening the black box of the operation and of observing the limits of this algorithmic moderation which is tending to become generalized.

> **The aim is to mark a critical pause in the face of the very powerful discourses that promote these technologies, or conversely the extremely critical discourses towards them.**

Indeed, their use is more complex. It must be considered in a more relational way with other entities. The governance dynamics of algorithms are entangled in a multitude of hybrid and collaborative modes of cognition. This governance is not stabilized and is undergoing many mutations that mix traditional methods with more sophisticated ones.

It is also a question of **considering the limits of their performance** (including in terrorist cases). We often have the illusion that everything is working well, especially as Facebook claims to have 98% efficiency in its quarterly transparency reports, as if they were going to solve everything, but it is much more complex. Clearly, the objective is to **go beyond the fetishization of algorithms**.

Therefore, the authors analyzed the role played by algorithms in the moderation of extremist content and saw how it was organized and used within a set of human and technical practices that mix different normative logics. Their objective was to be able to apprehend them within larger assemblies to understand their influences on the surveillance of information flows.

METHODOLOGY

The different **documentary sources and transparency reports**:

- Press articles
- Testimonies (e.g., Mark Zuckerberg in the Congress)
- Company blogs
- Terms of use

These documents were dated from **2016 to 2019**.

These platforms are increasingly being held to account by governments. There is enough data and information to map the different actors. The authors were able to trace the flaws in the algorithms that were highlighted by the platforms. There was enough information at this level, but it is necessary to keep a critical distance from these flaws because there are still grey areas. However, the companies are producing literature on this subject.

KEY FINDINGS

On the issue of algorithmic governance, what is clear is that in recent years both public and private actors have increasingly come to understand that the art of excluding hateful and terrorist content must be practiced through technologies, and more particularly through Artificial Intelligence. The issue of algorithmic moderation is becoming much more important in classifying and governing the content of large digital platforms. It has become a prominent feature.

What has been observed through this documentary analysis is that when we talk about algorithmic moderation (and here about extremist content), we see that we are not talking about a single algorithm but about **several algorithms that are involved in the moderation of content** and that are highly diversified and specialized. It is a **complex algorithmic ecosystem**.

Four main algorithmic functions have been identified:

1. **Image matching** compares new posts, videos or images posted to a base of existing content. By comparing with previously collected images in a database, the algorithm will remove an image posted by a user that matches. This is a manual process of collection and comparison.
2. **Content detection and classification technologies** that use machine learning, including natural language processing. The algorithm will generalize from known examples, learn, and then make classifications. These are more complex techniques that tend to become more widespread with technical progress.
3. **Terrorist cluster detection techniques** will look at users' networks to see if they are following other extremist users.
4. **Algorithms for detecting recidivist accounts**, i.e., accounts that have already been deleted and reappear on platforms.

The authors also find, based on the literature reviewed, that **algorithms report more content than humans**, delete more content and delete it faster. Algorithms find content by themselves and not by user reports. The platforms studied report very high-performance rates (98% for YouTube and Facebook, 93% for Twitter).

However, there are many **limitations to these algorithms**. First, they perform less well in understanding the context of speech. Furthermore, firms describe a difficulty in developing integrative technology that works on different types of media (e.g. on video and text). In addition, algorithms often focus on a particular terrorist group such as the Islamic State (formerly EI) and Al qaida but are sometimes unable to spot other groups due to differences in language and types of propaganda. In 2019, detection tools for hate content were only developed in 30 languages for hate speech and 19 languages for terrorist propaganda (figures to be verified).
Finally, it is difficult to regulate live and instantaneous content. For example, in the case of the Christchurch attack, the video lasted 20 minutes before being deleted. The algorithm was not able to train itself on this type of content, so it was slow to remove it directly. But progress has been made in identifying certain themes.

But the authors also studied **its relational character with other entities**. Indeed, with their ultra-high-performance figures of over 95%, these algorithms can be perceived as completely autonomous entities.
These technologies require experts, engineers, NGOs, civil society, governments, etc. The authors note that the more firms have developed these technologies, the more people they hire in their security and safety team.

Expertise for algorithms is needed in two main areas: in its design and in its operation. Algorithms can be poor regulators in complex contexts that require more interpretation. Platforms are therefore inviting users to participate in the moderation of content. There are now collectives and communities of Internet users who have taken on the mission of denouncing hateful content, accounts that promote extremism, etc. These communities are using discursive means to combat and eliminate hateful and extremist speech. These methods also foster an environment in which more lateral and horizontal surveillance flourishes.

Firms and platforms are increasing partnerships with other technology companies, governments, civil society groups, academics, and NGOs. The idea is always to foster shared learning about hate speech and propaganda mechanisms. These different communities can report pages, profiles, or share photos and videos associated with extremist groups to feed the platforms' databases. This content regulation embodies chains of association between humans and machines, experts, and non-experts.

The point of this article is that all these platform security measures order the goals of the new and the old with respect to surveillance. For example, they involve horizontal whistleblowing practices by humans and quite a few other security actors. The idea, when you look at the algorithms, is that you have lots of heterogeneous entities interacting. So, these findings somewhat erode the notion of the algorithm as a representative and primary protector of control.
Therefore, the authors use the notion of **cognitive assemblages**, developed by Katherine Hayles. Securing online content is involved in cognitive assemblages with a multitude of actors working together. Algorithmic temporality fits well with the speed of content sharing, but little with the diversity of the empirical world.

KEY TAKEAWAYS

- Algorithmic governance is much more fluid, cobbled-together and plural than one might think. It brings together many moving human and machine entities that mix traditional and more sophisticated methods. This assembly is not stable, it is moving. This action of observing and consolidating knowledge on profiles and contents at risk (which must be removed) is extremely distributed and collective.
- This governance is only partially algorithmic, contrary to a vision that fetishizes algorithms.
- The article does not aim to minimize the liberticidal character of these technologies, but it must be recognized that they are a cognitive assembly and that they are not in themselves the carriers of control. Everything operates in rather large cognitive assemblies.
- This knowledge on these assemblies is not stabilized, further studies are needed.

The Role of Consumer Perceptions of the Ethics of Machine Learning in the Appropriation of Artificial Intelligence-Based System

Christine Balagué, Zeling Zhong

Christine Balagué

Christine Balagué is HDR Professor at IMT-BS and holder of the Good in Tech Chair (www.goodintech.org). Her research focuses on modelling the behaviour of connected individuals, ethics of technology and AI, and responsible digital innovation. She is also a member of several national committees: CSA expert committee on online disinformation, Defence ethics committee, Haute Autorité de Santé recommendations impact commission, executive committee of Cap Digital. As Vice-President of the National Digital Council from 2013 to 2016, she is also co-author of several reports submitted to the French government on digital issues. She published more than 50 research articles in scientific journals or international conference proceedings as well as several books on society and economic digital metamorphosis.

Zeling Zhong

Zeling Zhong is a researcher in management in Good in Tech Chair. Her research focuses on technology appropriation and Internet of Things devices. She defended her dissertation at IMT-BS / University Paris Saclay in november 2019 on the topic: Understanding Smart Connected Objects Appropriation : a modelling approach using hierarchical components. Zeling Zhong's thesis won the ANDESE prize and was finalist of the AFM (Association Française du Marketing) thesis prize.

WHY IS THIS TOPIC IMPORTANT ?

The specificity of this article is to link ethics of artificial intelligence to consumer reaction. The main research question is: **does making algorithms fairer and more transparent improve consumers' appropriation of the technology?**

The contribution of this research is to analyze the consumer's perception of algorithms that are more transparent, more explainable, and therefore more ethical, and its consequences.

In this paper, the authors study **four dimensions of ethical artificial intelligence**: **data privacy and security, fairness, accountability, and transparency** of Machine Learning (ML) algorithms, and their impact on consumer's appropriation of Machine Learning systems.
The results show that the more the technology respects ethical criteria, the more the consumer will appropriate the technology. It is therefore a question of reconciling ethics and the market.

The subject is important because first, it is in line with the development of Trustworthy Artificial Intelligence, approach developed by several international reports since 2016 on the ethics of AI, based on the respect of principles (e.g. justice, autonomy, beneficence, non maleficence in the report on Ethics guidelines for Trustworthy AI from the High Level Expert Group of the European commission). Second, the consumer's perception of Trustworthy Artificial Intelligence remains a little-studied topic but essential for companies developing machine learning technologies or using embedded machine learning systems in their products and services. This paper highlights that trade-offs between ethics and algorithms performance impacts consumers' perception positively.

METHODOLOGY

The methodology relies on **Partial Least Squares (PLS) structural equations models** on latent variables (e.g., appropriation, measured by six dimensions : knowledge, consciousness, self-adaptation, control, creation, and psychological ownership; data ethics dimensions: algorithms' data privacy and security, fairness, accountability, and transparency).

The authors conduct a fieldwork on **a large database of 5 000 consumers** having virtual personal assistants (Google Home, Alexia, etc.) at home. The respondents answered a **quantitative questionnaire measuring consumers' perception of several latent variables**, in particular appropriation and ethics dimensions of the virtual assistant, i.e., perceived transparency, fairness, accountability, privacy.
They use PLS techniques to model AI services appropriation, as well as its antecedents (ML ethics and trust) and consequences (perceived value and NPS). The results show that ML ethics play a key role on consumer's behavior by positively influencing AI services appropriation.
The paper also reveals a positive effect of appropriation on product perceived value and Net Promoter Score (NPS), which is a classic index of product recommendation. The authors also show that there is a mediating effect of trust: the more it is perceived that the algorithms are transparent, fair, accountable, and respectful of privacy, the more consumers trust the technology and the better they will appropriate it.

KEY FINDINGS

There is a **positive effect of consumers' perceived ethics on appropriation of machine learning** technologies. The more consumers perceive transparency, fairness, accountability and data privacy of the AI technology, the more people will have confidence in the technology.

Artificial Intelligence appropriation can be measured by **six dimensions: knowledge, consciousness, self-adaptation, control, creation, and psychological ownership**.

KEY TAKEAWAYS

- The authors encourage companies to move towards the ethics of AI or trustworthy Artificial Intelligence as it is a brake neither on innovation nor on market.
- Marketers implementing AI based services must select more transparent, accountable, and fair algorithms in their technologies. Indeed, the respect of these criteria facilitate their AI services appropriation by customers.
- It is important to fullfil contractual ethical commitments because of its mediating role between user trust and ML ethics.

Where does the value of data lie?

Accepted and presented in Conference international 4S (Society for the SocialStudies of Science)

Kevin Mellet & Thomas Beauvisage

Kevin Mellet

Kevin Mellet is an Assistant Professor of Sociology at Sciences Po, and a researcher at the CSO. He is the Scientific head of the "Marketing & society" master's degree within the Innovation Management School of Sciences Po. His research work mainly draws on economic sociology and science & technology studies to study market techniques in the digital age. Current research interests focus on the emerging data marketing landscape and the formation and regulation of the personal data economy. His
involvement in the 'Good In Tech' Chair is twofold. First, he promotes and relays the Chair's activities and calls for projects within the community of Sciences Po teachers and researchers, in conjunction with the two chairholders, Dominique Cardon and Christine Balagué. Then, he is involved as a researcher in the activities of the chair, with a research project on the compliance practices of companies in the field of relationship marketing.

Thomas Beauvisage

Thomas Beauvisage is a sociologist and web scientist at the Social Sciences Department of Orange Labs (Sense). His PhD focused on web usage mining and browsing behavior characterization. His current research interests include online participatory market devices, advertising and e-reputation, and online media usage. He also collaborates on methodological research on the use of quantitative behavioral material for social sciences.

WHY IS THIS TOPIC IMPORTANT ?

This work comes from a set of questions related to the way in which, both in scientific studies and in the public and media, the issue of the data economy is addressed. **The idea that data is the oil of the 21st century, i.e., a resource on which to build innovation, growth, wealth, etc., is widespread today**. The authors note that data and value are immediately equated. Similarly, data would be an available resource that only needs to be extracted, like oil. This image is very present, and many works discuss it by affirming, for example, that data should not be framed as a commodity, but as a capital, a resource in which one must invest.

This raises a twofold question: **how do those who follow this line of reasoning construct their evaluations, their measures of what the data economy is? Is it obvious? Can we not deconstruct this evidence?**

The authors show that the large macro evaluations that seek to quantify the weight of the data economy (trillion-dollar valuations by BCG in 2012 and EU for 2025) are extremely fragile in their construction and their ability to evaluate what data produce. Data is everywhere and there is no real data market (or else it is the tip of an iceberg). Most of the value created by data is linked to the search for productivity gains, innovation, etc.

Very often research will tend to attribute to data the value produced by a set of actors (the Tech actors). For example, it is common to hear that if Google and Facebook are so powerful, it is because they have more data and process it with less sensitivity to privacy. This idea is found in Shoshana Zuboff's book The Age of Surveillance Capitalism. But this claim is largely exagerated when you look at the economic success of these companies.

Example: Google and Facebook business models

Data is not really involved in the formula that transformed both firms into nearly infinite cash machines. Google's success was built on its search engine and its ability to develop an advertising system adapted to its search engine and the queries people make. Similarly, Facebook's success is linked to breaking an initially given rule and introducing advertising into the newsfeed (users and their exchanges). When Facebook developed in 2011 on mobile, there was no space to put advertising on the sides. That's why they integrated advertising in the newsfeed and that's when their revenues skyrocketed. Finally, data comes late in this development and is not central, although it is obviously important today.

In the same way, there is some overstatement about the economic importance of players that we will call "data brokers" (Acxiom, Epsilon, Bluekai, etc). They certainly pose real problems, but we must consider that these "pure" data merchants are economic dwarfs when compared to Google or Facebook. In fact, what seems to work is the combination of data + computational capacity + network effects + attention.

A pragmatist look at valuation practices and situations

The authors propose a positive contribution by maintaining that it is necessary to look at the value of data with a pragmatist lens. Hence, one need to start from several local situations and configurations: **How, in a given situation, do economic actors and professionals integrate data into their work? And above all, how do they assign a value to these data?** Is it, then, possible, to estimate the contribution of this thing identified as “the data” to their work?

The authors argue that it is from these micro-observations that we can try to start again with a sounder analysis of the data economy and the value of data. Indeed, during the surveys they conducted, the authors realize that **it is not the data that produces the value** or that it is not as valuable as that.

Example: Google and Facebook business models

The advertiser has access to the Facebook Add manager, the interface that allows to set up advertising campaigns: defining a creative ad, campaign objectives, and the targeted audience, i.e., the targeting parameters. Here, the data will enable advertisers to find micro-segments, create very fine audiences, enrich audience segments, etc. But, when looking at the day-to-day work of advertisers who advertise on Facebook, when it comes to micro-targeting, the use of interest categories, these tools do not work well. These segments are expensive to create and tend to reach a sub-category of users who are often over-solicited, which limits the effectiveness of the tool. Advertisers will seek to reach rather large segments and not ultra-precise ones. The promise of personalization and hyper-targeting (thanks to data!) has its limits because it doesn't allow for a broad reach.
We are typically in a configuration in which the value of data is not obvious and must be weighed against other factors and players at play.

METHODOLOGY

The conference paper "Where does the value of data lie?" develops ideas in a theoretical manner but it is based on investigations that the authors have been able to carry out on different aspects of the economic and market use of data in the context of their work.

They investigated the actors who sought to measure the data economy: consulting firms such as BCG, international institutions (European Commission, World Economic Forum, International Monetary Fund), etc. The objective was to map these measurement initiatives and show their methodological weaknesses.

The authors also drew on past empirical surveys of professionals, particularly in the data marketing and advertising sectors, to understand how they integrate, question, and evaluate data in their daily work (media planners, advertising buyers, advertising agencies, etc.).

KEY FINDINGS

The authors' intention is not to dispute the fact that data are valuable. But the question "where does the value of data lie?" admits no easy answer.

1. This is important to handle the "data=value" equation with caution

2. The broad **macro-economic quantifications of the data economy**, especially those that compare data to oil, **are very fragile**.
3. Data are obviously very important and central because they are involved in many activities and areas but estimating their value means looking at the actual work of these professionals. Often, we observe that **the contribution of data is not so obvious**. This is an observation that can be made historically (the success of Facebook and Google), but also with advertisers for whom targeting is not necessarily relevant.

KEY TAKEAWAYS

The authors make two strategic recommendations:

- We must try to **overcome the wave of excitement**, enthusiasm, fantasy but also fear, surrounding data. The idea that Big Data is what we should invest in should be qualified. Instead, we need to start again from concrete situations, from daily work, to requalify and understand the value of data.

- We need to **pay attention to measurement issues**: what are we measuring and how are we measuring? In concrete terms, we need to evaluate the value increments produced by data, starting from micro situations, and from there estimate whether to invest. It is necessary not to fall into the trap of rushing into a preconceived notion that "data is value, it is necessarily the axis of growth n°1".

Issue 4 - April 2022

Good In Tech RESEARCH NEWS

Rethinking innovation and technology as drivers of a better world for and by humans

Christine Balagué
Professor at Institut Mines-Télécom Business School
Co-holder of the Good In Tech Chair
christine.balague@imt-bs.eu

Dominique Cardon
Professor at Sciences Po and director of the medialab
Co-holder of the Good In Tech Chair
dominique.cardon@sciencespo.fr

Jean-Marie John-Mathews
Data scientist
Coordinator of the Good In Tech Chair
jean-marie.john-mathews@imt-bs.eu

Jade Vergnes
Writer for Good In Tech Research News
jade.vergnes@sciencespo.fr

Clic here to contact