


The Crisis of Social Categories in the Age of AI

Jean-Marie John-Mathews*  a, b

Dominique Cardon^{a, c}

^a Sciences Po (France)

^b Université Paris-Saclay (France)

^c Université Paris-Est (France)


Submitted: December 20, 2022 – Revised version: January 23, 2023

Accepted: February 1, 2023 – Published: March 15, 2023

Abstract

This article explores the change in calculation methods induced by deep learning techniques. While more traditional static methods are based on well instituted categories to measure the social world, these categories are today denounced as a set of hardened and abstract conventions that are incapable of conveying the complexification of social life and the singularities of individuals. Today AI models try to overcome some criticism raised by rigid social categories by combining a “spatial and temporal expansion” of the data space, producing a global transformation of the calculation methods.

Keywords: AI; machine learning; social categories.

*  jeanmarie.johnmathews@sciencespo.fr

1 The Change in Calculation Methods of Our Societies

This article explores the recent shift in the way we calculate our societies (Cardon et al., 2018). The use of computational techniques to assist decisions is not new. Computational methods have long been used to rank and select individuals, with the help of a computer script, to verify a form's compliance with deterministic decision rules, for instance. With the arrival of machine learning (ML) tools, these methods provide a technological solution for decision-makers' confusion when dealing with their uncertainty faced with files combining an increasingly large list of entities and events. Decision-maker may feel indeed helpless faced with the diversity of points of reference offered by files for orienting around different principles. The quality of their decisions can much more easily be criticized (Hahn & Tetlock, 2005) on various grounds: they prioritized certain criteria over others; their social homogeneity hides structural biases; they were not attentive to the diversity of variables that could lead to other outcomes; and so on. Faced with disparate and extensive files, the introduction of automated tools to assist in decision-making based on machine learning models proposes the replacement of unstable justification of decisions with statistical probability. These tools order variables when the candidate comparison space becomes less comprehensible. The introduction of a statistical score is carried out today in a very different way, depending on the domain. It sometimes adopts the form of no more than an additional piece of information in the file, such as in the case of the prediction of the likelihood of a repeat offence in American judges' decisions to grant bail; or it can have a greater automaticity, such as to guide the police to places where crimes are more frequent (Brayne & Christin, 2020). As draft legislation on AI use shows, the question of the automaticity of results is one of regulators' main points of intervention in seeking to "keep a human in the loop" (Jobin et al., 2019). This article proposes to link this shift in calculation methods with the growing social criticism of statistical categories, what we call the crisis of social categories. We claim that this shift in decision-making toward ML has been favored by the incapacity of category-based methods (rules using criteria, etc.) to cover the variety and multiplicity of world events¹. Finally, we argue that this shift of the continuation of a more general spatio-temporal expansion of data space.

The appearance of machine learning techniques introduces a change in statistical culture that warrants some attention (Breiman, 2001). One of the particularities of these methods is that they are unaware of the decision rule in advance; they learn from the data. To set up this type of model, it is necessary to train an algorithm with a dataset (training database) composed of both input data (files) and the output results of previous decisions. The model is then adjusted by trial and error to make the prediction error based on training as small as possible (Goodfellow et al., 2016). If the model is learned based on the correspondence between input and output data, the decision rule can no longer be grounded on criteria-based justifications which are a priori stable and automatic. The model governing selection is a statistical approximation of the best way to compare file variables in relation to the given objective. A traditional way of presenting the operations governing the design of such a model consists in defining three separate spaces (Cornuéjols et al., 2018; Mitchell, 1997). Input data constitute the *observed space*, and the results of the calculation constitute the *decision space*. Between these two, the designer of the calculation must imagine a *hypothesis space* (sometimes also called the

1. In an upcoming article, we are going to develop more deeply the theoretical and sociological background of this claim. Using Boltanski's distinction between the reality and the world, we describe how this shift is linked with our institution's capacity to produce justifications and maintain the categories of reality against the criticism coming from the world.

“constructed space”, “latent space”, or “version space”). The latter space does not exist and cannot be empirically observed; it is an imaginary space in which the designer projects the ideal variables that should preside over the algorithm’s decision (Figure 1).

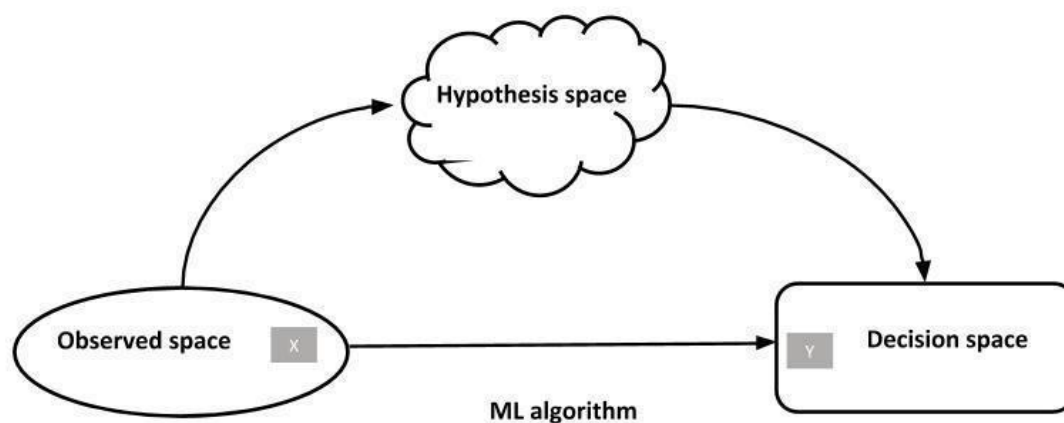


Figure 1. Categorical machine learning and the *hypothesis space*

For example, in the case of recruitment, the designer of a hiring algorithm sets the goal of selecting intelligent, motivated, and competent candidates. However, these ideal variables — which establish the basis of the justification of the algorithmic decision — do not exist in the data of the observed space. The hypothesis space contains ideal qualities that the designer wants to be the basis of the decision. Because it is impossible to have objective data allowing one to unequivocally incorporate these principles, it is necessary, in the observed space, to find good approximations so that the decision can effectively be made on the basis of the qualities projected in the hypothesis space. This is the task entrusted to the learning algorithm: discovering, in the observed data, the model that best approximates the ideal variables of the decision. In traditional uses of machine learning methods, the designer is asked to verify the quality of the relationship between the observed space and the hypothesis space. They must be attentive to the choice of data used in the learning model. They can use an input variable to verify that the relevant variables as a function of the goal have not excessively deformed the results with respect to this control variable. This is how we can measure demographic bias, for instance if the selected goal has contributed to systematically excluding certain populations. In machine learning, the idea that the designer is capable, at the very least through approximation, of controlling the semantic link between the observed space and the hypothesis space has long constituted the base point of the possibility to assess their fairness.

2 Social Criticism of Statistical Categorization

This change in calculation method is the consequence of a constant increase in the size of files. However, this increase in the number of data points also reflects the transformation of our societies’ relationship with forms of categorization of reality. With the individualization and complexification of social interdependencies, individuals’ criticism of the taxonomies into which they are categorized is being expressed more and more vehemently (Bruno et al., 2015). The regular patterns shown by social statistics no longer have an impact on our societies. They are

often denounced as a set of hardened and abstract conventions that are incapable of conveying the complexification of social life and the singularities of individuals (Desrosières, 2014). This movement is first and foremost evidenced in the effects of the criticism, developed in particular by the social sciences, aimed at revealing the constructed and artificial nature of statistical categorization (Desrosières & Thévenot, 1988; Boltanski, 2014). We are also witnessing a challenging of national indicators, such as the unemployment rate, consumer price index, or GDP, which are often identified as statistical constructions that can be manipulated according to the political constraints of the moment. The statistical pictures produced by government organizations are perceived as being reductive and have led, under the pressure of econometrics in particular, to profound transformations in statistical instruments that replace socio-professional categorization with continuous variables such as income (Angeletti, 2011). The production of categories also encompasses worldviews originating in the history of power relations in our societies which hide structural asymmetries between groups, such as patriarchy, institutional racism, or the economization of the world (Crawford & Calo, 2016; Eubanks, 2017; Murphy, 2017).

Social science's destabilization of broad statistical aggregates intersects with the development of increasingly harsh social criticism with respect to categorical representations of society, without the former being the cause of the latter. This criticism is based on our societies' individualization logics and the imperative of singularization that is becoming an increasingly heavy burden for individuals (Martuccelli, 2010). Individuals are expressing, more strongly than before, the desire for self-representation based on their chosen identities rather than being represented by the statistical categories assigned to them. At all crossroads of social life, people demand not to be reduced to the category that supposedly represents them. They refuse to allow themselves to be enclosed in the socio-professional categories that served to lay the foundations of a status-based society. Patients no longer wish to be reduced to their disease, customers to their purchases, tourists to their routes, militants to their organization, witnesses to silence, and so on. The classification systems, which became categories for perceiving the social world, struggle to operate as interpretation instruments shared by all. This *subjectivist* critique of categorical assignment first and foremost rejects the homogenizing nature of excessively large categories that suppress intra-categorical variability. Under the effect of the developing role of economics and social science as expertise in public and private organizations, an immense undertaking of refinement and granularization has taken place, densifying and increasing the accuracy of the measurements of statistical devices. Society is no longer simply broken down into strata related to social status, but rather displays inequalities between genders, geographic origins, religions, life paths, etc. The multiplication of the principles used to differentiate individuals thus contributes to decreasing the centrality of the opposition between social classes. This enables the emergence of new dimensions of individuals that had not been registered in the traditional statistical system, such as personal biography, network of acquaintances, mobility, etc. The comprehensibility of social structure has thus become more complex due to the multiplicity of dimensions according to which individuals can be compared to one another.

However, criticism of the categorical representation of society also encompasses a more radically subjectivist dimension. These representations of the social world are also criticized for their normalizing effect and their inability to take into account the personal and experienced dimensions of the feeling of discrimination. The representation of the social world developed by pragmatic sociology, for example, is not composed of fixed entities that can be described by variables with monotonic causal relationships with one another. In contrast with the idea of a "general linear reality" (Abbott, 1988), many disciplines within sociology believe that the mean-

ing of entities changes depending on the place, situation, or interactions. They also believe that they are very poorly depicted by the set categories of traditional statistical instruments. Deconstructed by social science, refused by individuals, imprecise for those producing them, and too rigid for comprehensive sociology, statistical categories are accused of poorly representing the singularity of individuals when subjected to tests.

3 The Expansion of the Data Space

Computational techniques to assist decisions closely combine a spatial and temporal expansion of the candidate data space, producing a transformation of the calculation method underlying decisions.

3.1 The spatial expansion of the data space

While one of the responses to criticism of statistical categories is to increase the number of data points, the dynamic triggered by the digitalization of the calculation and of data is taking a new turn: the displacement of *criteria* with *variables* is being combined with a transformation of variables into *traces*. An example can serve to illustrate this new shift. Consider two databases used in the literature on loan allocation in the machine learning field. The first is the German loan database (Hofmann, 1990). It contains data on 20 variables for 1,000 former loan applicants. Each applicant was categorized as “good credit” (700 cases) or “bad credit” (300 cases). The majority of the variables are qualitative, such as civil status, gender, credit history, the purpose of the loan, or work, and can contain several encoded categories. For example, variables related to employment are encoded with 5 possible attributes: unemployed, non-qualified employee, qualified employee, manager, and highly qualified employee. This set of traditional data is characterized by the fact that it contains few variables, that each variable is subdivided into several modalities, and that these modalities provide a stable and homogenous description of each data point.

The second dataset is a typical example of the new type of databases that the digital revolution has made possible. Like the first, it was used to generate credit scores by Lu et al. (2019) in an article that won the prize for best article at the ICIS 2019. The main feature of this dataset is that it is composed of a much larger number of data points and that the number of variables increased spectacularly. For each loan applicant, the platform collects personal data such as: (i) records of online purchases (in other words, the order time, product name, price, quantity, type of product, and information on the receiver); (ii) mobile telephone records (in other words, call history, mobile telephone use, detailed use of mobile applications, GPS mobility trajectories); and (iii) social media use (in other words, if the borrower has an account, (if so) all of the messages posted with time stamps and presence on social media, including number of fans, follows, comments received, and “likes” received on weibo.com). Each variable is no longer broken down into a small number of modalities but rather into a series of granular pieces of information: shopping list, telephone calls, geolocated travel, behaviour on social media. The data are no longer aggregated as variables but rather resemble information flows that are as elementary as possible. The first dataset is composed of categorical entities, whereas the second is composed of granular flows of behavioural traces. The first seeks to convey, in the candidate comparison space, personal states that can be inferred to be related to a loan application; the second does not make this effort, considering that all information available represents some-

thing from the candidate's world, without, however, being able to relate it to the candidate's capacity to repay a loan.

This initial data transformation process is at the heart of the promise of big data. By expanding the network of information, files adopt an unprecedented form and enable the establishment of *continuous data* that is supposed to be much closer to the candidate's world. Instead of sampling interpretable variables, the calculator breaks down information into a set of flows that extend into the candidate's past as well as that of other candidates whose future is known, assimilating increasingly larger portions of individuals' lives and behaviour (Rouvroy & Berns, 2013; Lury & Day, 2019). *Continuous databases* record a priori a set of individuals' behaviours, choices, and actions in order to insert them into a comparison space, the scope of which expands to candidates' worlds. The data characterizing them no longer associate them with a candidate's identity but rather seek to record and transform something more continuous in their life path. This transformation in the format of data has been identified by many analyses. It establishes, alongside individuals, a "data double" (Haggerty & Ericson, 2000, p. 611; Lyon, 2001; Krasmann, 2020), an "informational person" (Koopman, 2019), a "data dossier" (Solove, 2004), or a "doppelganger" (Harcourt, 2015) which constitutes a person's digital shadow by totalizing a fragmented set of behavioural information. In reality, these data are more difficult to unite in a single point or to articulate to one another, and as it is often much more difficult to extract relevant predictions than the privacy panic feeding much research in the domain could imagine (Salganik et al., 2020). It nevertheless constitutes a new identity production regime, a few features of which can be highlighted.

The first feature is that the digital framework used to collect this information — however pervasive, distributed, and increasingly discrete and opaque — nonetheless remains fragmented and extremely specific to a particular type of activity (movement, using your debit card, surfing the web, etc.). Moreover, it is largely controlled by the private operators of large digital services (Mau, 2019). These *traces* come to exist only because a network of sensors has been implemented, in other words, a metrology for collecting and archiving these signals and the artefacts used to calculate them. The spread of these probe and sensor devices in our societies, the fact that they are spanning increasingly more numerous and varied territories are now interpreted as a tipping point towards a new form of capitalism (Zuboff, 2019).

The second feature is related to the fact that this regime of knowledge on individuals can be described as "post-demographic" (Rogers, 2009). These data longer to seek to capture individuals in the traditional formats of social statistics (gender, age, education, profession, etc.). They define not a stable state of identity but rather an extensive and disconnected set of oscillating and partial *micro-states* describing a conduct, a location, a feeling, or an expression (Terranova, 2004). The granular fragmentation of these signals adopting the form of a trace that detaches a specific activity from an individual's course of action is often compared to the concept of "dividuals" introduced by par Gilles Deleuze (1990). This transition from variables to traces has significant consequences for the ways in which individuals are statistically associated with a whole. By virtue of the conventions of equivalency underpinning large-scale statistical categorizations, any categorical attribution can be related to a population of which the categorized individual is one case within a known distribution (Le Bras, 2000). Categorical identification allows us to relate the results of a categorization to a more global distribution, and therefore to shine light on inequality or unfairness in test results. However, when an individual's trace is compared to a set of other traces which can no longer be compared with one another in terms of a given category, revealing unfairness becomes much more difficult (John-Mathews et al., 2023). By eliminating the categorization of individuals, the test loses its reference population,

which in turn makes it more difficult to show that the test results are unfair.

A third feature is related to the behavioural nature of these continuous signals, which seek to capture individuals in a way that is as close as possible to their living state and biology. This discards the possibility of individuals' reflexive self-identification with the symbolic forms representing them. These new databases (purchasing behaviour, browsing history, behaviour on social media, location, sensor recordings while driving a vehicle, success or failure during a multiple-choice test on a MOOC video, etc.) do not record explicit performances carried out in specialized arenas but rather turn each course of action into a micro-performance. In this sense, it is not an exaggeration to consider social life, at any point at which it encounters a sensor, to be a theatre for testing individuals and an infra-political space that remains partially opaque and invisible. As John Cheney-Lippold (2017) demonstrates in his analysis of this "soft biopolitics", this continuous data generates statistical aggregates which produce "measurable types" with no symbolic referents in the categorical grammar of identities. Google is completely indifferent to whether a user identifies as "male", "female", or another non-binary label; all it needs is probabilities estimating the behaviour of this user as being 74% "female". Through trial and error, the algorithm tests different trace aggregates in such a way as to get as close as possible to the goal. The calculation itself condenses aggregates of meaning that are interrelated by variable and changing predictive paths. Last of all, these paths between the observed space and the *decision space* move away from social actors' definition of their identity and therefore lose intelligibility. The trajectory of the decision from the observed space to the decision space follows a series of changing branches, passing through aggregates of signals forming sorts of chimaera (patterns) to which it is very difficult to assign meanings. Nonetheless, novel attempts do now exist in the XAI (explainable artificial intelligence) literature, to try to view these paths within the hidden layers of neural networks, such as the work of Chris Olah at OpenAI.²

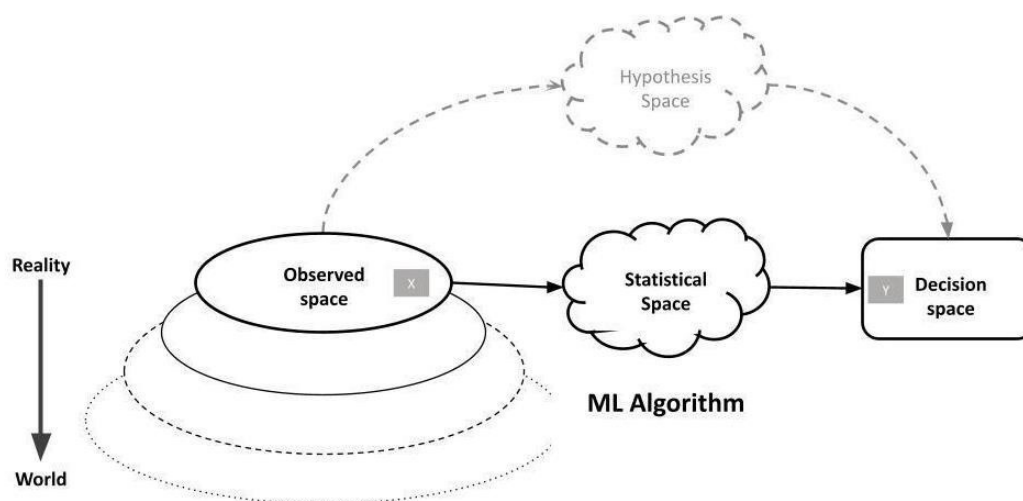


Figure 2: The hypothesis space makes way to a statistical space

When the number of input variables increases significantly when they adopt the form of unlabeled traces, it becomes impossible to project a hypothesis space and to comprehend the relationship between the observed space and the goal of the machine learning. In this displacement

2. <https://distill.pub/>

(Figure 2), the hypothesis space darkens until becoming a “black box” that it is increasingly difficult to elucidate, even though a great deal of highly innovative IT and algorithm ethics research is seeking to overcome this difficulty (Olah et al., 2017). The hypotheses that we can make regarding the variables are based no longer on input data, but on the goal that serves to adjust the model. The hypothesis space, as both an ontological and normative externality, no longer provides the possibility of the expression of criticism. This displacement in the calculation architecture makes it difficult a reformulation of the ethical issues of machine learning, such as explicability or fairness of results (John-Mathews, 2021; John-Mathews, 2022).

3.2 The temporal expansion of the data space

The second direction that this expansion process follows is temporal. It is characterized by a reorganization of the temporalities of the calculation transporting the modeling of files toward a future prediction (Cevolini & Esposito, 2020). The goal used to adjust the model is no longer immediately associated with the actual data of the file but is projected at a future expectation that becomes increasingly more distant from the data gathered by the test.

Consider another simple example. The most basic learning models set a temporal variable directly associated with the temporality of a test as their goal: namely, passing or failing. Candidates who will pass the test have been designated by a model that has learned to recognize them, based on a database composed of the files of previous winners. However, the hypotheses formed to justify the variable used as the goal of the learning never define present qualities exclusively, but also powers to act in the future. Successful candidates are chosen because they are assumed to have the skills that will allow them to execute a given type of performance in the future. Juries, judges, or recruiters constantly make these types of hypotheses during their deliberations, without having the possibility of objectifying them in the form of a numerical probability. The expansion of the comparison space enabled by quantification techniques promises to provide an estimate of this future probability using the traces of former test candidates. Their current performances become the goal for modeling the selection of new candidates.

The data used to produce the learning model therefore incorporates information that goes beyond the temporal context of the test to evaluate future performances; For instance, the ranking of candidates for a heart transplant is henceforth guided by the prediction of life expectancy in good health following the transplant (Hénin, 2021); the hiring of salespeople at a company is guided by an estimate of its turnover five years from now; and the inspection of the technical systems of boats is determined by a prediction of having an accident at sea. Judicial risk evaluation instruments seek to estimate, prior to a trial, whether a defendant is a threat to public safety, or if he or she is at risk of repeat offending should bail be granted (Harcourt, 2006). This projection of the calculation towards the anticipation of future performance has the goal of pushing back the boundaries of the unknown by incorporating potentialities within a space that can be measured by probabilities.

4 Conclusions

The boundaries of this new configuration of data call for a discussion that is outside the scope of this article. However, we believe that it is important to explore the way in which the evaluation of individuals' qualities is now increasingly dependent on socio-technical systems. From this point of view, a critical analysis opportunity for social science can be established, consisting of a much more closely investigating of how the calculations of neural networks work. Announc-

ing the disappearance of the hypothesis space — as we have done — is simply a way of saying that a calculation's intelligibility has become so complex and strange that social science should abandon its goal of understanding forms of sociotechnical life. A more novel and ambitious approach would be to truly pay attention to the diversity of the proposals and paths traced by these new types of calculations within information spaces to which we are not accustomed.

References

- Abbott, A. (1988). Transcending General Linear Reality. *Sociological Theory*, 6(2), 169–186. <https://doi.org/10.2307/202114>
- Angeletti, T. (2011). Faire la réalité ou s'y faire. La modélisation et les déplacements de la politique économique au tournant des années 70. *Politix*, 3(95), 47–72. <https://doi.org/10.3917/pox.095.0047>
- Boltanski, L. (2014). Quelle statistique pour quelle critique?. In I. Bruno, E. Didier & J. Prévieux (Eds.), *Stat-activisme. Comment lutter avec des nombres?*. Paris: Zones.
- Brayne, S., & Christin, A. (2020). Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts. *Social Problems*, 68(3), 608–624. <https://doi.org/10.1093/socpro/spaa004>
- Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, 16(3), 199–215. <https://doi.org/10.1214/ss/1009213726>
- Bruno, I., Didier, E., & Prévieux, J. (Eds.). (2015). *Statactivisme: Comment lutter avec des nombres*. Paris: Zones.
- Cardon, D., Cointet, J.-P., Mazières, A. (2018). Neurons Spike Back. The Invention of Inductive Machines and the Artificial Intelligence Controversy. *Réseaux*, 5(211), 173–220. <https://doi.org/10.3917/res.211.0173>
- Cevolini, A., & Esposito, E. (2020). From Pool to Profile: Social Consequences of Algorithmic Prediction in Insurance. *Big Data & Society*, 7(2). <https://doi.org/10.1177/2053951720939228>
- Cheney-Lippold, J. (2017). *We are Data. Algorithms and the Making of our Digital Selves*. New York, NY: New York University Press. <https://doi.org/10.2307/j.ctt1gk0941>
- Cornuéjols, A., Miclet, L. & Barra, V. (2018). *Apprentissage Artificiel: Deep Learning, Concepts et Algorithmes*. Paris: Eyrolles
- Crawford, K., & Calo, R. (2016). There Is a Blind Spot in AI Research. *Nature*, 538, 311–313. <https://doi.org/10.1038/538311a>
- Deleuze, G. (1990). Post-scriptum sur les sociétés de contrôle. *L'autre journal*, 1.
- Desrosières, A. (2014). *Prouver et gouverner. Une analyse politique des statistiques publiques*. Paris: La découverte. <https://doi.org/10.3917/dec.desro.2014.01>
- Desrosières, A., & Thévenot, L. (1988). *Les catégories socioprofessionnelles*. Paris: La découverte.

- Eubanks, V. (2017). *Automating Inequality, How High-tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press. <http://www.deeplearningbook.org>
- Haggerty, K.D, & Ericson, R.V. (2000). The Surveillant Assemblage. *British Journal of Sociology*, 51(4), 605–622. <https://doi.org/10.1080/00071310020015280>
- Hahn, R.W., & Tetlock, P.C. (2005). Using Information Markets to Improve Public Decision Making. *Harvard Journal of Law and Public Policy*. 29(1), 213–289.
- Harcourt, B. (2015). *Exposed. Desire and Disobedience in the Digital Age*. Cambridge, MA: Harvard University Press. <https://doi.org/10.4159/9780674915077>
- Harcourt, B. (2006). *Against Prediction: Profiling, Policing and Punishing in an Actuarial Age*. Chicago, IL: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226315997.001.0001>
- Hénin, C. (2021). Confier une décision vitale à une machine. *Réseaux*, 225(1), 187–213. <https://doi.org/10.3917/res.225.0187>
- Hofmann, H.J. (1990). Die Anwendung des CART-Verfahrens zur statistischen Bonitätsanalyse von Konsumentenkrediten. *Zeitschrift für Betriebswirtschaft*, 60, 941—962.
- John-Mathews, J.-M. (2021). Critical Empirical Study on Black-box Explanations in AI. *arXiv*, 2109.15067. <https://doi.org/10.48550/arXiv.2109.15067>
- John-Mathews, J.-M. (2022). Some Critical and Ethical Perspectives on the Empirical Turn of AI Interpretability. *Technological Forecasting and Social Change*, 174, 121209. <https://doi.org/10.1016/j.techfore.2021.121209>
- John-Mathews, J.-M., Cardon, D., & Balagué, C. (2022). From Reality to World. A Critical Perspective on AI Fairness. *Journal of Business Ethics*, 178, 945–959. <https://doi.org/10.1007/s10551-022-05055-8>
- John-Mathews, J.-M., De Mourat, R., De Ricci, M., & Crépel, M. (2023). Re-enacting Machine Learning Practices to Inquire into the Moral Issues They Pose. *Convergence*, forthcoming.
- Koopman, C. (2019). *How We Became Our Data. A Genealogy of the Informational Person*. Chicago: The University of Chicago Press.
- Krasmann, S. (2020). The Logic of the Surface: On the Epistemology of Algorithms in Times of Big Data. *Information, Communication & Society*, 23(14), 2096–2109. <https://doi.org/10.1080/1369118X.2020.1726986>
- Le Bras, H. (2000). *Naissance de la mortalité. L'origine politique de la statistique et de la démographie*. Paris: Seuil/Gallimard.
- Lu, T., Zhang, Y., & Li, B. (2019). The Value of Alternative Data in Credit Risk Prediction: Evidence from a Large Field Experiment, ICIS 2019 Conference, Munich, December.
- Lury, C., & Day, S. (2019). Algorithmic Personalization as a Mode of Individuation. *Theory, Culture & Society*, 36(2), 17–37. <https://doi.org/10.1177/0263276418818888>

- Lyon, D. (2001). *Surveillance Society: Monitoring Everyday Life*. Buckingham: Open University Press.
- Martuccelli D. (2010). *La société singulariste*. Paris: éd. Armand Colin.
- Mau, S. (2019). *The Metric Society. On the Quantification of the Social*. Cambridge, MA: Polity Press.
- Mitchell, T. (1997). *Machine Learning*. Boston, MA: McGraw-Hill.
- Murphy, M. (2017). *The Economization of Life*. Durham, NC: Duke University Press. <https://doi.org/10.1515/9780822373216>
- Olah, C., Mordvintsev, A., & Schubert, L. (2017). Feature Visualization. *Distill*, 2(11). <https://doi.org/10.23915/distill.00007>
- Rogers, R. (2009). Post-Demographic Machines. In A. Dekker & A. Wolfsberger (Eds.), *Walled Garden* (pp. 344–355). Amsterdam: Virtual Platform.
- Rouvroy, A., & Berns, T. (2013). Gouvernamentalité algorithmique et perspectives d'émancipation. *Réseaux*, 1, pp. 163–196. <https://doi.org/10.3917/res.177.0163>
- Rouvroy, A., & Berns, T. (2013). Algorithmic governmentality and prospects of emancipation. Disparateness as a precondition for individuation through relationships?. *Réseaux*, 177(1), 163–196. <https://doi.org/10.3917/res.177.0163>
- Salganik, M.J., Lundberg, I., Kindel, A.T., Ahearn, C.E., Al-Ghoneim, K., Almaatouq, A., Altschul, D.M., Brand, J.E., Carnegie, N.B., Compton, R.J., Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B.J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., ... McLanahan, S. (2020). Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 117(15), 8398–8403. <https://doi.org/10.1073/pnas.1915006117>
- Solove, D. (2004). *The Digital Person: Technology and Privacy in the Information Age*. New York, NY: New York University Press.
- Terranova, T. (2004). *Network Culture: Politics for the Information Age*. London: Pluto.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power, Public Affairs*. London, UK: Profile Books.

Jean-Marie John-Mathews – Sciences Po (France); Université Paris-Saclay (France)

📄 <https://orcid.org/0000-0001-8049-1222>

✉ jeanmarie.johnmathews@sciencespo.fr; 🌐 <https://www.jeanmarie-johnmathews.com/>

Jean-Marie John-Mathews is a researcher in algorithmic ethics. He is the coordinator of the *Good In tech* academic chair (Institut Mines Télécom / Sciences Po) and teaches at Sciences Po, Université PSL and Université Paris-Dauphine. His research focuses on the development of so-called “responsible” tools in artificial intelligence and has been published in several scientific journals such as *Technological Forecasting and Social Change* and *Journal of Business Ethics* as well as in international conferences such as the *International Conference on Information Systems*.

Dominique Cardon – Sciences Po (France); Université Paris-Est (France)

🌐 <https://medialab.sciencespo.fr/en/people/dominique-cardon/>

Dominique Cardon is Professor of Sociology at Sciences Po where he directs the Médialab. His work focuses on the uses of the Internet and the transformations of the digital public space. His recent research focuses on Internet social networks, forms of online identity, amateur self-production, the analysis of forms of cooperation and governance in large online collectives, and the analysis of the algorithms used to organize information on the web. He has notably published *La démocratie Internet* (Paris, Seuil/La République des idées, 2010), *A quoi rêvent les algorithmes. Nos vies à l'heure des big data* (Paris, Seuil/République des idées, 2015) and, with Jean-Philippe Heurtin, *Chorégrapheur la générosité. Le Téléthon, le don et la critique* (Paris, Économica, 2016), *Culture numérique* (Paris, Presses de Sciences Po, 2019).