Laboratoire d'Analyses et Modélisation de Systèmes pour l'Aide à la Décision
UMR 7243

CAHIER DU LAMSADE

397

November 2020

Social Responsibility of Algorithms: an overview

Alexis Tsoukiàs

Social Responsibility of Algorithms: an overview

Alexis Tsoukiàs,
CNRS-LAMSADE, PSL, Université Paris Dauphine

Abstract

Should we be concerned by the massive use of devices and algorithms which automatically handle an increasing number of everyday activities within our societies? The paper makes a short overview of the scientific investigation around this topic, showing that the development, existence and use of such autonomous artifacts is much older than the recent interest in machine learning monopolised artificial intelligence. We then categorise the impact of using such artifacts to the whole process of data collection, structuring, manipulation as well as in recommendation and decision making. The suggested framework allows to identify a number of challenges for the whole community of decision analysts, both researchers and practitioners.

1 Motivations

There is increasing concern around us about the impact of using automatic devices making decisions for several aspects of our life, including credit scoring, admissions to Universities, pricing of goods, recommender systems, up to automatic vehicles or predictive justice (see [2], [24], [26]). However, the use of algorithms in order to automatise decision making is not recent ([13]); actually algorithms exist even before computer science became the industry we know. We can summarise the situation today under the following observations:

- We are creating and using autonomous artifacts with increasing decision autonomy.
- We have autonomous artifacts with increasing learning capacities.
- There is evidence of biased decisions, of counterintuitive decisions, of inappropriate use of personal and sensible data, of unforeseen consequences, when such devices are largely adopted and used[1].
- Software editing and data services are concentrated to few industrial players.

The aim of this paper is to clarify a number of issues which affect both researchers and practitioners interested in decision support (decision analysts). It turns out that many of the concerns we are discussing today, already existed in the literature (see for instance [52]) and are less "new" and "urgent" from what they appear to be. On the other hand, the extension today of designing, testing and actually using autonomous artifacts represents a real challenge for the community of decision analysts. The paper aims at identifying which are these challenges and how can we appropriately handle them.

The paper is structured as follows. In Section 2 we make a brief survey of the literature with no pretention to be exhaustive, essentially in order to identify the principal trends. Section 3 introduces the principal concepts through which we can establish a common framework. Section 4 presents two brief examples which help understanding the topics discusses in th previous section. Finally, Section 5 summarises the challenges we have in front of us the next years.

[1]The best known controversy is the "COMPASS" case: `https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing`

2 Historical background

The literature about Decision Support Systems dates back to the 70s: see the seminal paper [18] and the two well known books [25] and [47]. This literature builds upon already existing research and practice with "Management Information Systems" (see [32]). The idea is simple: exploit the information existing and circulating within an organisation in order to improve decision making under different types of requirements (see also the interesting discussion in [27]).

In more recent days the same idea came alone under the concept of "analytics" (or business analytics or business intelligence; see [12]). The "new" idea is to extend the use of data in order to support decision making creating and assessing massive data bases (more or less open), thanks to a large increase of computing capacity. However, the application of these ideas remains bounded at supporting "human decision makers" within organisations, the scope of "analytics" being to produce suitable information for decision makers.

A relatively innovative idea has been instead to increase the decision capacity of "autonomous artifacts" in order to improve the overall performance of complex systems. However, once again automatising decisions is not a totally new idea; we can see how this evolved through the following topics.

- Automatically conducted vehicles have been designed since a century ago: automatic pilots for aircrafts date at the beginning of the 20th century (see [1] or [48]). Automatically controlled devices and robots exist since the middle of the 20th century ([22], [50]) and represent today a very important scientific and industrial area.
- Multi-Agent systems started being designed in the 80s (see [44] or [53]) allowing software agents to perform with increasing decision autonomy.
- Recommender Systems appeared soon after as software platforms where consumers could be automatically guided among huge catalogs of goods and being advised about their choices matching their preferences with products features and the behaviour of similar consumers (see [3] or [42]).
- Blockchains introduced the possibility to decentralise trust construction procedures through distributed cryptography on the web (see [34], [35]).

As can be noted the idea of increasing the decision capacity of autonomous artifacts already has several decades of development, including

commercial and industrial applications of large scale (virtually any aircraft today is automatically driven and most e-commerce platforms include a recommender system). There have been though two breakthroughs:
- the increasing availability and accessibility to data (of any type and quality, including personal and sensible ones);
- the massive expansion of "deep learning algorithms" allowing high level correlations among data with excellent accuracy and predictive capacity (for a presentation see [17], while for an interesting discussion about correlation and causality see [37]).

Such developments fuelled a literature about the impact and the consequences of automated decision making. This literature has been essentially focussed around three areas.

1. **Fairness**. Since the seminal paper [14], there have been several tentatives in order to establish a general definition of "fairness" for decisions taken by algorithms. This notion of fairness assumes the existence of "protected" groups within the society, which are potentially threaten by biased algorithmic decision making processes (see also [21] and [28]). However, such "protected groups" are only recognised within certain countries and it soon appeared that there are several formal and substantial difficulties in establishing a model of general validity (see [15]).
2. **Accountability and Explicability**. Not independent from the fairness issue there has been the discussion about the accountability of algorithms (see [8] and [51]). The issue here is the possibility to provide convincing explanations on why an algorithm would end taking a certain type of decisions (possibly unfair, biased or counterintuitive). The topic includes explicability of data mining and machine learning algorithms (see [20]) with specific emphasis to the case where the algorithms behave as black boxes with unpredictable behaviour (such as deep neural networks).
3. **Ethics**. Finally there has been discussion about the ethical dimension of automated decision making. The issue arises essentially in the case of automatically conducted and/or unmanned vehicles and devices which may need to take decisions with high ethical impacts (such as impacting human life: see [5] and [39]). The topic however, has gone beyond this specific area questioning the possibility and/or opportunity to endow algorithms with ethical principles (see for instance [19]).

The result of such discussions has been the creation of new scientific communities, possibly interdisciplinary ones, the largest for the moment being the ACM-FAccT series of conferences (see `https://facctconference.org/index.html`).

3 What is the problem?

The survey presented in the previous section, far from being exhaustive, highlights the fast growth of an area of scientific investigation, but also of public concern. In reality there exist several different problems which both scientific paper and press and blogs tend to put together under different "titles" basically sharing a number of keywords: Artificial Intelligence, Data Protection, Algorithmic Transparency etc. (see [29], [36]). Most of them tend to raise concerns of the general public of how such technologies could impact our life. It pays, however to clarify a number of issues starting with establishing precisely the object of scientific investigation.

From our perspective this object is the *"design, implementation and systematic use of autonomous artifacts with enhanced decision capacity*. In the following we are going to analyse which are the different problems this object includes.

In conducting our analysis we will adopt an industrial production perspective because we are talking about the evolution of an industry whose raw material are data. Under such a perspective we are going to focus upon the raw material itself (the data), the transformation process (the algorithms), the implementation (the software), the outcome and the impact to the society. However, before analysing the components of this industrial process we may analyse a number of fundamental topics.

3.1 Fundamentals

The first fundamental topic to remember is that automation is not a straightforward perspective, but a choice. There are plenty of examples around us of processes which are not automated and nobody thinks to automatise them. If automation is a choice then there is somebody who makes the choice and there should be reasons for which this choice has been done. Automation for certain types of production has been decided by the industry and their management essentially in order to increase profits (although several times quality of the products has been used as a reason). Automation of certain industrial processes has been decided for safety purposes or in order to alleviate workers from unhealthy or dangerous activities. If automatising a decision process is a choice, we should always ask ourselves who decides to automatise, for which reasons and who is going to pay the cost of it. If the process to automatise concerns the public (such as college admissions or predictive justice) there is a matter of democracy and citizens' participation to such decisions.

The second fundamental topic to remember is that decisions imply responsibility and responsibility implies liability for the consequences of any decision. Each time we consider automatising a decision process we should always ask ourselves who is liable for the decisions taken by the autonomous artifact we create. In the flying industry this issue has been long time solved: liable are the airlines who use aircrafts with automatic pilot and there is a chain of responsibilities, certifications and training in order to keep such liability clear. The liability issue does not concern solely the principle, but also the practical aspect: be sure that responsibilities can be traced, recognised and affected to those who could be liable. Automatising a decision process means that we have a clear idea of how the liability issue is going to be considered.

The third fundamental topic concerns the fact that algorithms can mirror how our societies are, but cannot change them. It is clearly annoying discover through what an algorithms learns that our societies are unfair, discriminate minorities, behave aggressively, in other terms are politically incorrect. But these are the societies as our democracies shaped them. If we do not like them, there are democratic paths for changing our societies, but algorithms will always mirror what our societies actually are. We cannot introduce innovation in society just designing innovative algorithms.

3.2 The raw material

The raw material of the type of processes we are concerned are data. Data are collected, stored, retrieved and manipulated and each single step of these processes could have an impact upon the whole decision process to automatise. There are two basic topics to consider as far as the use (term resuming all the above steps) of data is concerned.

The first topic concerns the rights an individual (a citizen) and/or a group have upon certain data. Data (of any type) do not belong specifically to somebody and for certain types of data we could consider them as "commons". However, we can have certain rights upon certain data and as soon as these rights are established we can consider whether these can be traded. However, trading rights implies establishing clear contracts. The problem today is that there is an absolute information asymmetry (see [30]) between each single citizen and his rights on the one side and the data industry on the other side. Besides, there is a value scaling about data availability: the value of the rights I have upon my personal data alone is an extremely small fraction of the value of owing the rights of millions of individuals.

The second topic concerns the certification of the data used within automated decision processes. Biased data will result in biased outcomes. Noisy data will result in bad quality outcome. Corrupted data will result in unverifiable outcomes. There is necessity to certify the whole pipeline of collecting, storing and retrieving data used for any automated decision process (see [9]).

3.3 The outcome

First of all we need to make an important distinction. Autonomous artifacts can provide two types of outcomes: "decisions" and "recommendations". For this purpose we may define a decision as an *irreversible allocation of resources to tasks or actions*. In the first case is the artifact that makes such an allocation which results in some action being undertaken, while in the second case the artifact only makes a recommendation (generally to a human agent) which "decides".

From a practical point of view there are very few autonomous artifacts which actually have full decision autonomy and generally this concerns "low level" actions in automatic controlled devices (such as in self-conducted vehicles). Most of the automated decision processes concern in reality artifacts which suggest a certain action to be undertaken. It can be the case of credit scoring, of predictive justice scores, college admissions, job candidates screening etc.. However, this "final decision freedom" of the human agent is far from being a warranty about the controllability of the final outcome. Most automatically formulated recommendations are rarely contested and usually are followed by the human decision makers, which essentially explains why such *suggestions* are regularly considered as *decisions*. In the following we will focus on automated "recommendation" processes, since these are the most frequent (and complex).

A first issue to consider is the fact that the result of information manipulation is never straightforward: there is no (and will never exist) universal procedure through which we can obtain from raw data a synthesis. Data manipulation ought being *meaningful* (see [43]), *useful* (see [6]) and *legitimate* (see [49]), these requirements still allow for plenty different procedures. It is a matter of choice for the designers and users.

The second issue, following from the previous one, is that we may desire adding further properties to the outcome: we may desire having a recommendation which is *fair, unbiased, neutral etc.*. The fact is that there is no unique definition to such concepts. Both economists in mechanism design theory ([23], [31]) and computer scientists more recently ([15]) realised that there are several different ways to define notions such as "fairness", each corresponding to different hypotheses about the society, the inequalities within the society and the ways to prevent or to correct them. This means we need to establish both the requirement of a feature to meet and a formal definition for each requirement and how to test it.

Establishing which requirements the recommendations needs to meet is a matter of choice. The third issue is to know who decides which requirements an outcome of a given autonomous artifact have to be satisfied. Several of such requirements might be inconsistent one with respect to another. Somebody (who?) has to make a choice resulting in satisfying a certain property and thus, failing to satisfying another one. Under such a perspective it is important when designing an autonomous artifact to know which properties/requirements/axioms an automated recommendation procedure satisfies and which not. This is rarely the case today (the reader can check that no recommender systems specifies how notes are aggregated among users and products and thus nobody knows which properties are satisfied by such procedures).

What happens in case the autonomous artifact is "data driven": in other terms the outcome depends essentially upon the data feeding process, but the data manipulation is unknown (as happens for many black-box automated procedures)? The fourth issue related to the quality of the outcome concerns the "hidden values" embedded within many autonomous artifacts. Decisions and recommendations are never based directly on raw data. Between these and any decision there are "'preferences" or "values" which allow to compute a "choice" (or whatever else a decision or a recommendation may mean; see [10]). Preferences and values are always subjective and represent an individual or a society of individuals. If an autonomous artifact is able to make a decision or to compute a recommendation it means that somebody embedded within the artifact his/her preferences. And these are independent from how the artifact turns to learn out from the data feeding it. It turns out that is of paramount importance to know how values are actually embedded in any of such systems and/or how these are learned (see [16]).

3.4 The process

It is often the case that not only the outcome of a process matters, but also the process itself. This is both the case for automated decisions and automated recommendations. The former might need to be explained, justified, tested and proven to be "correct" in case of accidents, misbehaviour, unforeseeable consequences etc. The latter might need to be trusted, defended, argued, recused, might need to be convincing, trustworthy, understandable, etc.. In all such cases we need to check whether the autonomous artifact is *accountable*. However, there are several different levels of *accountability*.

1. Given an algorithm or to be more precise a bundle of algorithms setting an automatic decision procedure can we trace precisely what these algorithms do?
2. Provided that we can trace the execution of the algorithms, can we provide "explanations" (interpretable, understandable, usable) to any type of stakeholder about the choices done and the obtained results?
3. Provided we can trace and explain the behaviour of an algorithm can we provide the "ultimate reasons" for which the algorithm/automatic device made a precise decision or recommendation? If it is the case can we replicate the decision providing the same input?
4. Supposing the algorithm cannot guarantee replicability (for instance in case the algorithm learns each time is executed we cannot guarantee that for a given input the output will remain the same) what type of explanations/justifications/reasons would be considered satisfying in case of a dispute?

Besides the above introduced aspects of accountability there are also long term consequences to take into account when a certain type of autonomous artifact is largely used in the real world. How should we define accountability for the long term impact of e-commerce platforms using recommender systems (using certain types of algorithms) for promoting their business?

3.5 The implementation

Autonomous artifacts are essentially software. Certainly in the case of robots and other autonomous devices there are physical parts which are equally important, but the essential of what we are talking is software. Indeed algorithms and procedures not necessarily are implemented in software, but we are concerned with the ones who actually are used under form of computer programs.

The first issue to consider is the formal verification that a given software implementation of a bundle of algorithms endowing an autonomous artifact, actually does what these algorithms are expected to do. This is far from being self-evident and the more complex the artifact is, the more difficult the verification becomes.

The second issue concerns security. Any software implementation can be attacked and/or manipulated. We can certainly choose safer, redundant and highly protected implementations (as the stakes of the artifact scope increase; see the case of e-voting), but this comes at a price which needs to be commensurable to the benefits and the value of the automation.

The third issue concerns the use of open source software. While this apparently could be inconsistent with straight security requirements, open source software remains the ultimate possibility to analyse why an autonomous artifact actually acts as observed. While security issues can easily be handled even when using open source code, being able to check the code through collective intelligence processes remains a fundamental warranty for most accountability issues.

3.6 The impact to society

Introducing a drug to a living system has expected and unexpected consequences. It is exactly for this reason that new drugs before being cleared and allowed to be used are extensively tested under rigid protocols, are permanently checked and submitted to scrutiny and possibly can be retired from commerce. Usually there is an independent authority which takes care of this process. We are going to use this "metaphor" (introducing a new drug to a living system) in order to consider the long term impact of introducing an autonomous artifact in handling some aspect of our everyday life.

Should we demand a certification for the whole process (raw material, outcome, process and implementation) before allowing to release an autonomous artifact in the society? Should we create an independent agency or authority for this purpose? While many of the known unknowns can be handled through appropriate design and preliminary testing, the only way to discover the unknown unknowns is to do extensive testing and monitoring.

The use of autonomous artifacts for some types of business implies modifying the business model of the enterprise and/or the organisation introducing this "innovation". The issue is whether, stakeholders, consumers, users are aware of the consequences such a modification will introduce upon the goods and services delivered by that type of business. Organisational studies are plenty of innovation failure cases ([41]) for businesses, organisations and markets, because the process was poorly designed, not understood, not fitting the expectations, undesired etc. and this includes the choices about automatising decisions and processes.

The industry of automated decisions and recommendations is dominated by few big players, both with respect to the collection, storage, retrieval and use of data and services and with respect to software editing and engineering. Monopolies never benefited consumers and this case will not be an exception. This market, as many others, needs regulations and these need to be global.

If we need to audit autonomous artifacts and monitor their long term impact and if we need to establish global rules for this market we need to bear in mind that the life cycle of these products can be short (even very short) compared to the length of audits and regulations. It might make sense to be innovative as far as the timing of regulation is concerned if we do not want to miss any real opportunity to control.

4 Examples

The following examples are voluntarily not among the typical ones used within the Artificial Intelligence literature just in order to show that several issues discussed in this paper are far beyond the AI challenges.

4.1 Automatic pricing

Automatic pricing became popular since the late 70s because of the innovation introduced by American Airlines: yield management. The simple idea consists in adjusting prices and seats offered on commercial flights following the demand prediction and possible capacity of the airline (see [46]). Today is a regular practice, not only for the airlines industry: many retailers practice automatic pricing in order to optimise revenue management. That said, we can make a number of observations.

1. Implementing the automatisation of this activity has been a choice, both for profit maximisation purposes and for gaining competitive advantages for the first runners. It is less obvious whether this resulted in better and less cheap services for the customers. In any case it was not a choice of the users who can find exactly the same product at several, significantly different, prices (see [33]).
2. There exist several different economic models helping to compute automatically prices, depending upon the type of product, the type of the market and the hypotheses about the consumers' behaviour (see [11]).

It is actually unknown whether the choice of any among such models has been discussed before using them.

3. We know instead that adopting a precise model of pricing, considering the density of the competition (on the retailers market), can lead to unforeseen consequences, as in the famous "Stapler" case[2], where the same object (a stapler) was sold at different prices in different neighbours. The use of competition density resulted in discounting the object in the "rich neighbours" (high density) and sell it at full price in the "poor neighbours" (low density). Not necessarily this was the policy and will of the retailer.
4. Automatic pricing strongly depends upon the quality and timing of the necessary data feeding the economic and decision models of yield management. However, there is no warranty that the data pipeline for any of the retailers adopting such tools is reliable and trustworthy. This is all the more important in case part of the data feed a black-box learning procedure for which we may have no convincing justifications available. On the other hand, the liability for any "wrong" decision remains internal to the retailer who will just have to absorb the consequences in their business.
5. Automatic pricing modified how the travel industry is organised and influenced how people travel and organise their leisure time. In other terms it had a huge impact upon the whole society (as often happens when industries introduce new products or services). On the other hand what is the long term impact of such new patterns of mobility and leisure time consumption? Are these sustainable at a long run? Nobody ever discussed that, when yield management models and algorithms have been introduced (more than 40 years ago).

4.2 Voting

Voting is not an automatic decision procedure, or at least is not perceived as such. However, the reader will note that when we adopt the term voting we implicitly consider voting procedures (algorithms) which "compute" a "winner" of an electoral contest. There exist several such algorithms and generally they may yield totally different results even when the preferences of the voting society are clear. The fact is that we (the society) need to make choices of which among such algorithmes should be used and possibly we

[2] https://www.wsj.com/articles/SB10001424127887323777204578\
\189391813881534

(the society) have to trust that the result is legitimate. For a presentation of electoral systems see [40], while a theoretical investigation about social choice theory can be found in [4] or in [45]. For an interesting survey about such methods being considered under a computational aspect see [7]. Once again we make a number of observations.

1. We vote in order to elect representatives, presidents, committees, chairs etc. and this is done for legitimating governing. However, "electing" is not the only way to appoint representatives, committees, chairs ... Our societies (after centuries of struggles) decided to use such procedures (which might result in less efficient decision procedures, but certainly more legitimated). We vote because we want to.
2. As already mentioned there exist many different voting procedures and algorithms computing the winner(s). It is well known that it does not exist and it will never exist an universal procedure, because even simple "democratic" requirements are inconsistent among them and cannot be satisfied simultaneously. This means we need to choose one. In doing that it pays knowing which requirements are satisfied and which not and this has been the scope of large part of the social choice theory literature.
3. Different voting procedures promote different views about our societies, the ways to govern the society and about citizens' participation (see [38]). Moreover, each of such systems need to make choices on how "fair representation" should be interpreted (proportionality among citizens, among regions, among ethnic groups are typical topics which are typically impossible to satisfy all together). These are political choices which need to be discussed as such and not as technical problems.
4. Electronic vote is increasingly popular, but has been tested to be easily hacked, corrupted and manipulated, while manual procedures are far more complicated to alter (at least under usual democratic institutions operating). As already noted previously, the software version of an algorithm does not coincide with the algorithm itself.

5 Conclusion

Let us try to summarise our overview. Does it make any sense to talk about the *social responsibility of algorithms*? Technically speaking, no, since algorithms cannot be liable for what they compute. Designers, clients demanding for algorithms, software editors can be considered responsible (and thus, liable), but not the algorithms. On the other hand, the use of algorithms in order to improve our decision making is older that computer science itself and the demand for extending their use, for creating further autonomous artifacts with decision capacity is never lasting.

As decision analysts we share part of the responsibility of how such autonomous artifacts are shaped, designed, implemented and used in the real world. Under such a perspective we should pay attention and further develop our theoretical research as well as our reflection about our practices around the following topics:
- characterising algorithms and procedures through the properties they satisfy or do not satisfy;
- remembering that each time we choose a precise procedure in order to solve a given decision problem this is rarely a straightforward choice, but one among many options and as such needs to be justified and considered for its impact beyond that precise problem;
- characterising and specifying the data to be used by algorithms, reflecting the three basic requirements: meaningfulness, usefulness and legitimacy;
- analyse how our methods, procedures and protocols are used and adopted within real organisations and within our societies.

Algorithms and formal models will never stop being used in order to improve how decisions are taken both by humans and machines. It is upon the designers define what improvement means and for whom. This is our social responsibility.

Acknowledgements

This document follows discussions which took place during the two workshops about this topic in Paris, December 2017 and 2019 (`www.lamsade.dauphine.fr/sra`) and I am indebted to the participants for their contributions. Although several ideas are due to these discussions, I remain the sole responsible for this essay.

References

[1] K.J. Åström and R.M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, Princeton, 2008.

[2] D. Abu-Elyounes. Contextual fairness: A legal and policy analysis of algorithmic fairness. *Journal of Law, Technology and Policy*, 2020:1 – 54, 2020.

[3] Ch.C. Aggarwal. *Recommender Systems: The Textbook*. Springer Verlag, Berlin, 2016.

[4] K.J. Arrow. *Social choice and individual values*. J. Wiley, New York, 1951. 2nd edition, 1963.

[5] J.-F. Bonnefon, A. Shariff, and I. Rahwan. The social dilemma of autonomous vehicles. *Science*, 352(6293):1573 – 1576, 2016.

[6] D. Bouyssou, T. Marchant, M. Pirlot, P. Perny, A. Tsoukiàs, and Ph. Vincke. *Evaluation and decision models: a critical perspective*. Kluwer Academic, Dordrecht, 2000.

[7] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A.D. Procaccia. *Handbook of Computational Social Choice*. Cambridge University Press, Cambridge, 2016.

[8] C. Casteluccia and D. Le Métayer. Understanding algorithmic decision making: opportunities and challenges. EPRS study, European Parlament, 2019. 104 pages.

[9] V. Christophides, V. Efthymiou, Th. Palpanas, G. Papadakis, and K. Stefanidis. End-to-end entity resolution for big data: A survey. Technical report, ARXIV, 2020. https://arxiv.org/pdf/1905.06397.pdf.

[10] A. Colorni and A. Tsoukiàs. What is a decision problem? preliminary statements. In *Proceedings of ADT 2013, LNAI 8176*, pages 139 – 153. Springer Verlag, Berlin, 2013.

[11] R. Cross. *Revenue Management: Hard-Core Tactics for Market Domination*. Broadway Books, New York, 1997.

[12] Th.H. Davenport and J.G. Harris. *Competing on Analytics: The New Science of Winning*. Harvard Business School Press, Harvard, 2007.

[13] Th.H. Davenport and J.G.Harris. Automated decision making comes of age. *MIT Sloan Management Review*, 46:11, 2005.

[14] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. Fairness through awareness. In *Proceedings of ITCS'12*, pages 214–226, 2012.

[15] S.A. Friedler, C. Scheidegger, and S. Venkatasubramanian. On the (im)possibility of fairness. Technical report, CoRR abs/1609.07236, 2016.

[16] J. Fürnkranz and E. Hüllermeier. *Preference Learning*. Springer Verlag, Berlin, 2010.

[17] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, Boston, 2016.

[18] G.A. Gory and M.S.Scott-Morton. A framework for management information systems. *MIT Sloan Management Review*, 13:55 – 70, 1971.

[19] J. Greene, F. Rossi, J. Tasioulas, K.B. Venable, and B. Williams. Embedding ethical principles in collective decision support systems. In *Proceedings of the AAAI 2016*, pages 4147 – 4151, 2016.

[20] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):93, 2018.

[21] S. Hajian and J. Domingo-Ferrer. A methodology for direct and indirect discrimination prevention in data mining. *IEEE transactions on knowledge and data engineering*, 25:1445 – 1459, 2013.

[22] V.D. Hunt. *Smart Robots: A Handbook of Intelligent Robotic Systems*. Chapman and Hall, London, 1985.

[23] L. Hurwicz and S. Reiter. *Designing Economic Mechanisms*. Cambridge University Press, Cambridge, 2006.

[24] D. Keats Citron. Technological due process. *Washington University Law Review*, 85:1249 – 1313, 2008.

[25] P.G.W. Keen and M.S. Scott-Morton. *Decision Support Systems: an Organizational Perspective*. Addison Wesley, Reading, 1978.

[26] J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson, and H. Yu. Accountable algorithms. *University of Pennsylvania Law Review*, 165:633 – 705, 2017.

[27] M. Landry, D. Pascot, and D. Briolat. Can DSS evolve without changing our view of the concept of problem? *Decision Support Systems*, 1:25–36, 1983.

[28] B. Lepri, N. Oliver, E. Letouzé, A. Pentland, and P. Vinck. Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy and Technology*, 31:611 – 627, 2018.

[29] B. Lepri, J. Staiano, D. Sangokoya, E. Letouzé, and N. Oliver. The tyranny of data? the bright and dark sides of data-driven decision-making for social good. In Cerquitelli T., Quercia D., and Pasquale F., editors, *Transparent Data Mining for Big and Small Data. Studies in Big Data, vol 32*, pages 3–24. Springer Verlag, 2017.

[30] A. Mas-Colell, M.D. Whinston, and J.R. Green. *Microeconomic Theory*. Oxford University Press, New York, 1995.

[31] E. Maskin. How to Implement Social Goals Mechanism Design. *American Economic Review*, 98:567–576, 2008.

[32] R.O. Mason and I.I. Mitroff. A program for research on management information systems. *Management Science*, 19:475–487, 1973.

[33] A.G. Mauri. Yield management and perception of fairness in the hotel business. *International Review of Economics*, 54:284 – 293, 2007.

[34] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. Technical report, bitcoin.org, 2008.

[35] A. Narayanan, J. Bonneau, E. Felten, A. Miller, and S. Goldfeder. *Bitcoin and cryptocurrency technologies: a comprehensive introduction*. Princeton University Press, Princeton, 2016.

[36] C. O'Neil. *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown, New York, 2016.

[37] J. Pearl. *Causality*. Cambridge University Press, Cambridge, 2009.

[38] D.W. Rae. *The Political Consequences of Electoral Laws*. Yale University Press, Yale, 1971.

[39] I. Rahwan. Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*, 20:5 – 14, 2020.

[40] A. Reynolds, B. Reill, and A. Ellis. *Electoral system Design. The new International IDEA Handbook*. IDEA, Stockholm, 2005.

[41] K. Rhaiem and N. Amara. Learning from innovation failures: a systematic review of the literature and research agenda. *Review of Managerial Science*, in press, 2019. https://doi.org/10.1007/s11846-019-00339-2.

[42] F. Ricci, L. Rokach, B. Shapira, and P.B. Kantor. *Recommender Systems Handbook*. Springer Verlag, Berlin, 2011.

[43] F.S. Roberts. *Measurement theory, with applications to Decision Making, Utility and the Social Sciences*. Addison-Wesley, Boston, 1979.

[44] S. Russel and P. Norvig. *Artificial Intelligence: a modern approach*. Prentice Hall, New York, 1995.

[45] A.K. Sen. Social choice theory. In K.J. Arrow and M.D. Intriligator, editors, *Handbook of mathematical economics*, volume 3, pages 1073–1181. North-Holland, Amsterdam, 1986.

[46] B.C. Smith, J.F. Leimkuhler, and R.M. Darrow. Yield management at American Airlines. *Journal on Applied Analytics*, 22:8 – 31, 1992.

[47] R.M Sprague and E.D Carlson. *Building Effective Decision Support Systems*. Prentice Hall, Englewood Cliffs, 1982.

[48] B. Stevens and F. Lewis. *Aircraft Control and Simulation*. Wiley, New York, 1992.

[49] A. Tsoukiàs. On the concept of decision aiding process. *Annals of Operations Research*, 154:3 – 27, 2007.

[50] N. Wiener. *Cybernetics*. MIT Press, Cambridge, 1948.

[51] M. Wieringa. What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In *Proceedings of FAT 2020*, pages 1 – 18, 2020.

[52] T. Winograd and F. Flores. *Understanding Computers and Cognition*. Ablex, Norwood, 1986.

[53] M. Wooldridge. *An Introduction to Multi-Agent Systems*. Wiley, New York, 2002.