Accountability in Artificial Intelligence[1]

Dr. Olga Gil
olgagil@ucm.es
Universidad Complutense de Madrid
Instituto Complutense de Ciencias de la Administración (ICCA)

September 8th, 2022

Abstract

This work stresses the importance of AI accountability to citizens and explores how a fourth independent government branch/institutions could be endowed to ensure that algorithms in today´s democracies convene to the principles of Constitutions. The purpose of this fourth branch of government in modern democracies could be to enshrine accountability of artificial intelligence development, including software-enabled technologies, and the implementation of policies based on big data within a wider democratic regime context. The work draws on Philosophy of Science, Political Theory (Ethics and Ideas), as well as concepts derived from the study of democracy (responsibility and accountability) to make a theoretical analysis of what artificial intelligence (AI) means for the governance of society and what are the limitations of such type of AI governance. The discussion shows that human ideas, as cement of societies, make it problematic to enshrine governance of artificial intelligence into the world of devices. In ethical grounds, the work stresses an existing trade off between greater and faster advancement of technology, or innovation on the one hand, and human well being on the oher, where the later is not automatically guaranteed by default. This trade off is yet unresolved. The work contends that features of AI offer an opportunity to revise government priorities from a multilevel perspective, from the local to the upper levels.

Keywords

artificial intelligence, accountability, democracy, regulation, philosophy of science, ethics, political theory

[1] Please quote as: Gil, Olga, 2022. "Accountability in Artificial Intelligence" Paper presented to the XVI Congress of the Spanish Association of Political Science, Girona, Spain, September 8th, 2022. Preprint.

I. Introduction

Throughout human history, one of main sources of social and political development have been new ideas, and the capacity of human beings to challenge the status quo through the conceptualization and application of these new ideas to upcoming challenges. Nowadays, artificial intelligence, including software-enabled technologies and big data are defended as a new contribution to governance. These technologies are powerful tools for smart government and smart city development (Kitchin, 2016). As much as these technologies may be an aidé for governance, problems with artificial intelligence are manifold. One of them is related to accountability.

This work reviews the above problem using tools ranging from philosophy, political theory, public administration and political science and argues that 1) Ideas about the future are very relevant to evolving societies, and thus, artificial intelligence, including software, and data driven technologies per se are not utmost mechanisms to make the choices that citizens and governments shall make, and 2) the context for participation and accountability within societies are issues to take into account when artificial intelligence development, including software technologies, and data driven policies are addressed. Artificial intelligence development, overarching software-enabled technologies, and big data as attributes of smart government development cannot subsede the wider contexts and issues in which these technologies are convened. All in all, the work argues artificial intelligence may help to make informed decisions on smart government development. However, it might not replace human agency in the policy agenda.

The work shows that ideas, as cement of societies, make it problematic to enshrine governance just into the world of devices, such as an artificial intelligence that includes either software-enabled technologies and big data. The work also shows that features of artificial intelligence make it an opportunity to revise government priorities from a multilevel perspective, from the local to the upper levels. In ethical grounds, the work stresses an existing trade off today between pushing for greater and faster advancement of technology, and human well being, where the latter is not automatically guaranteed by default. This trade off could be solved through a positive win-win game which is yet unresolved both at the current stage of research and also at the practical and applied development of public and private governance instruments today. Finally, it proposes a fourth independent government branch endowed to ensure that algorithms in today's democracies convene to the principles of Constitutions. The purpose of this fourth branch of government in modern democracies would be to enshrine accountability of artificial intelligence development, including software-enabled technologies, and the implementation of policies based on big data within a wider democratic regime context.

II. Definitions

This work argues that we may refer to software-enabled technologies, big data and algorithms as parts of artificial intelligence, as new developments in science might include software-enabled technologies, and big data into the same construct. An algorithm has been defined as a mathematical construct with a "finite, abstract, effective, compound control structure, imperatively given accomplishing a given purpose under given provisions (Hill, 2016)". Algorithms might be automated or human run, and are used to make sense of the data.

When we turn to big data, it is interesting to analyze the perspectives of scientists, from engineering to informatics. These perspectives would fall into definitions of legitimacy that Matheson (1987) would include in his work about *Weber and the classifications of forms of legitimacy* as legitimacy based on expertise. Among these experts, however, there are critics of big data even as mythology, where an aura of truth, objectivity, and accuracy is assumed by default.

The work of Mikalef et al. (2018) includes a systematic review of the literature on big data from this legitimacy based on expertise perspective, and it shows several streams of definitions. In those definitions there is a ladder of complexity, from those focusing on quantity and management, to those looking at the general context in which data is extracted from. We start with the first steps of complexity in the definition of big data. For Russom (2011), Bekmamedova and Shanks (2014), and Davis (2014) Big Data includes the data storage, management, analysis, and visualization of very large and complex datasets. White (2011), besides large volumes of data, includes new analytical technologies and business possibilities, such as sensor data, web and social media data, improved analytical capabilities, operational business intelligence and cloud computing. Supporting big data would combine these technologies. For Beyer and Laney (2012), for Schroeck et al. (2012), and for Sun et al. (2015) big data involves high-volume, high-velocity, and/or high-variety information assets together with new forms of processing to enable enhanced decision making, insight discovery, and process optimization. This definition is also within the scope of Bharadwaj et al. (2013), for whom big data refers to datasets with sizes beyond the ability of common software tools to capture, curate, manage, and process the data within a specified elapsed time. For Gantz and Reinsel (2012), Big data focuses on three main characteristics: the data itself, the analytics of the data, and presentation of the results of the analytics.

Stepping up in complexity there is another stream of definitions, that includes societal problem solving as a main end of the tool. Within this stream, we find Kamioka and Tapanainen (2014), for whom Big data is large-scale data with various sources and structures that cannot be processed by conventional methods and it is intended for organizational or societal problem solving. Here we also find Boyd and Crawford (2012) making a wider analysis, including context. For Boyd and Crawford big data is a cultural, technological, and scholarly phenomenon that rests on the interplay of (1) Technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large datasets. (2)

Analysis: drawing on large datasets to identify patterns in order to make economic, social, technical, and legal claims. (3) Mythology: including the widespread belief that large datasets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy. From a critical perspective as well Constantiou and Kallinikos (2015) raise concerns about big data severe problems of semiotic translation and meaning compatibility.

Reading Stamboliev (2019) we may amplify the scope of the problem on accountability to robots, and those would include humanoid companions, dataveillance, and algorithms. Stamboliev defends that and morality and ethics shall be linked to include all of them, specially in relation to populations such as the elderly and vulnerable groups, and from a democratic perspective we could include other minority groups. This brief introduction on definitions also shows the constant need for reconceptualization about what artificial intelligence shall include as technologies evolve.

III. The governance of artificial intelligence: accountability and ethics

There are two scientific streams covering consequences of the use of artificial intelligence and big data, including public and private governance. The first one, represented by Floridi and Taddeo (2016) focuses on segmentation of morality into attributes, discussed metaphorically. Deontological codes, consent, privacy of data subjects and secondary use would be part of an ethic of practices, as a segment of morality. Stamboliev, instead, proposes an approach that opens the door for political theory and other disciplines such as institutional law, public administration and regulation. In this view, accountability is linked to practicall responsibility over legal and policy conflicts (Stamboliev, 2019). Especially important, according to Stamboliev in a context of robotics being a network of individuals programming, developing, and deciding, where "we need discussions on practical accountability" (Stamboliev, 2019). These two approaches, in essence, show the existence of a high level of contestation about concepts regarding algorithms and how to relate them to human control.

From an epistemological point of view and the perspective of philosophy of science, Gallie (1956) sets out an analytic framework consisting of seven criteria for essentially contested concepts --where we could include **artificial intelligence**. These criteria have been reviewed by Mulligan, Koopman, and Doty (2016, p. 5) as follows:

— Appraisiveness. The concept must signify or accredit certain valued achievement.
— Internal complexity. While the concept's 'worth is attributed to it as a whole' [14, p. 172], it must be a 'normative concept[s] with internal complexity' [15, p. 150]; it must also be multi-dimensional, with different principles of operation, values to be served, objectives and justifications.
— Diverse describability. Owing to the internal complexity, disputants can describe the concept in different ways, using different features and according to them different weight.
— Openness. The concept has to allow for unpredictable and unprescribed changes over time to address evolving circumstances.

— Reciprocal recognition. Conceptions of the concept have to be used and maintained against other uses both aggressively and defensively. Gallie's claim requires that disputants must recognize the contestation and have some sense of the underlying criteria at the source of the disagreement with others; however, many have questioned whether mutual recognition is required.
— Exemplars. The competing conceptions of the concept must derive from an original authoritative exemplar (generally thought to allow for multiple, rather than a singular, paradigmatic example) acknowledged by all disputants.
— Progressive competition.The continuous contestation must contribute to sustaining and/or developing the concept in an optimum manner.

Beside appraisiveness, internal complexity, diverse describability, openness, reciprocal recognition, and exemplars, this work argues that progressive competition that includes philosophy and the social sciences is needed to analyze accountability of artificial intelligence.

This work has explained in previous sections that a contested concept is an unresolved issue pertaining to artificial intelligence. We now turn to stress epistemological issues related to science in general and artificial intelligence in particular (Ongaro, 2019). The following table points out three epistemological problems that humanity and science face today, namely, that we live in an epoch in which old certainties crumble, that truth is not guaranteed through method, and that value based judgements are key in decisions both in public and private sectors.

Table 1: Definition of philosophy, and unresolved issues related to philosophy and science.

	Philosophy. Definition and and today´s main issues
DEFINITION “Philosophy is concerned with the acquisition of rational knowledge and understanding of reality in its entirety ...its purpose ...is deepened understanding of reality, rather than vaster disciplinary knowledge of a subject matter, like in individual scientific disciplines”. In the words of the ancient Greek philosophy is “love of wisdom” (Ongaro, 2019, 137).	Philosophy, too important to be sidelined. Ultimate purpose, the betterment of public governance, public policies and public services
	1) Philosophical wisdom is of utmost value to the training of student, researchers, experts and practitioners
	2) We live in an epoch in which old certainties crumble
	3) There is a need to check whether knowledge is tipped towards the technical side, to the detriment of the conceptual tools to decipher a more complex world: "There is a lot of theory... mostly of very technical level... ticking disciplines into the false imagining that one has captured the world formula... Much talk about evidence, by what actually is meant empirical proof, is an indication of such a position, as the idea of an openness of data that somehow guarantees truth through method" (Ongaro, 2019, 137)
	4) Need to address issues of value based judgements enabling the making of decisions in executive roles both in private and public administrations

As this work introduces, in a world and an epoch in which old certainties crumble, philosophy helps to make the argument that artificial intelligence, including software development and data driven technologies might become central elements to make decisions about the polity and society. Bringing philosophy in, the betterment of public governance, public policies and public services, and even regime type, have proved collective human endeavors that today might not be subsumed to data collection and algorithms.

IV. Floridi and Taddeo versus Stamboliev. Debates on ownership, literacy, accountability and programmers

We have overviewed concept contestation in sciences, and notice the interest on progressive competition. We have also noticed epistemological limitations on concepts related to artificial intelligence. We follow on, discussing progressive competition of the artificial intelligence concepts under different approaches. First of all approaches focusing on morality, and secondly, an approach stressing accountability.

In the first place, resting on morality, the analysis of Floridi and Taddeo (2016). Ethics of data science, under this approach, includes ethics of data, ethics of algorithms and ethics of practices (Floridi and Taddeo, 2016). Whitin ethics of data, we would group privacy (reidentification and group privacy), trust in whom, and transparency (of what). Within the ethics of algorithms, following Floridi and Taddeo (2016), we would find responsibility and accountability, the ethical design of algorithms requirements, and the ethical auditing of algorithms. Within the ethic of practices, we would find deontological codes, consent, privacy of data subjects and secondary use of data.

Departing from ethics and morality as basis to public auditing of algorithms is Stamboliev work (2019): “I advocate to establish more clarity on data ownership, privacy, and agent literacy, and on the potential commodification of data when robots are applied.” (Stamboliev, 2019, 249) When Eugenia Stamboliev writes about data ownership she is introducing and linking three concepts: ownership, literacy (Gil et al. 2015) and accountability. Stamboliev defends that accountability shall be linked to practical responsibility, an aspect attached to legal and policy conflicts around robots or tracking systems. This view opposes a focus on a segmentation of morality into attributes, which are then discussed metaphorically, as Floridi and Taddeo propose (Stamboliev, 2019, 254-255). Stamboliev's proposal is that “In the context of robotics being a network of individuals programming, developing, and deciding, we need discussions on practical accountability as soon as possible (2019, 255).” This practical discussion shifts away from making robots, or artificial intelligences, more ethical and it proposes, instead, to focus on programmers:

> *“I urge (...) to shift away from only focusing on how to make robots more moral or ethical and to examine the developer and programmers’ ambitions in applying or designing robots. I did not imply to look for a stable and fixed human entity or a rational subject who owns stable moral values, but I indicated that the human agent (as individual researcher or as collective industry) is (or must be) the decisive authority in the wider context of using robots” (Stamboliev, 2008, 256).*

Stamboliev concludes that “a common denominator of research should be to aim for progress that supports human well being, instead of pushing for a greater and faster advancement of technology” (Stamboliev 2019, 257).

An additional issue comes with data collection and who collects data,

> *“Robotic devices have not yet played a major role (...) but this will change once they are more common outside industry. Robotic devices together with the ‘Internet of things’, the so-called ‘smart’ systems (phone, TV, oven, lamp, home, ...), the ‘smart city’ (Sennett 2018) and ‘smart governance’, are set to become part of the data-gathering machinery that offers more detailed data, of different types, in real time, with ever more information – to whomever has access” (Müller, 2019, 5).*

When we think further about accountability, how accountability might take place, and how responsibility can be allocated, As Müller (2019) recalls, “With distributed agency comes distributed responsibility.” (Taddeo and Floridi 2018, 751). How this distribution might occur is not a problem that is specific to AI, but it gains particular urgency in this context (Nyholm 2018), e.g. for autonomous vehicles. In classical control engineering, distributed control is often achieved through a control hierarchy plus control loops across these hierarchies” (Müller, 2019, 14).

Accountability focuses on the aims of policies and the policy process and the role of citizens on it. Increasing citizen accountability is the proposal of Shahrokni, H., & Solacolu, A. (2015), who defend a technology enabled paradigm based in information and communication technologies (ICT). ICT would enable citizens in everyday decision-making processes, which in turn could result into a transition of increased citizen responsibility that could entail an unprecedented number of daily decisions and ethical trade-offs. Thus, the argument goes, software enabled technologies and in big data shall go hand in hand with a model of citizenship and polity development in which citizens are enabled and empowered to participate in everyday decisions, including those related to ethical issues.

Value based judgements are also linked to the existence -or the absence- of a strategy to address the desired development of artificial intelligence in a modern society, including software-enabled technologies, such as robots and big data for smart government. This strategy, that shall be addressed from any smart government perspective, might be lacking in plans that overview the future. Taking the example of the report ‘Preparing for the Future of Artificial Intelligence in the United States and others released by the White House Office of Science and Technology Policy (OSTP) reports emphasize increased economic prosperity, improved educational opportunity, social security and quality of life, and enhanced national and homeland security. However, there is a lack of revision of the open possibilities to renovate a social pact in which democracy rests: “although important, these issues are approached in a way that can best be summarized as trying to fit artificial intelligence (AI) into the specific vision of US national priorities, instead of seeing the new features of AI as a good opportunity to revisit these priorities, both nationally and internationally” (Cath et al., 2018). The development of strategies to address the future of artificial intelligence in our societies brings us to the question of human ideas and sources of change.

V. Ideas about the future and uncertainty as a source of change

Understanding artificial intelligence as an opportunity to revise government priorities from multilevel perspectives allows us to turn now to ideas as a focal point of future societies. Ideas about the future have been very relevant to evolving societies, and it follows that artificial intelligence, including software and data driven technologies might not be the ultimate mechanism for the choices that citizens and governments shall make. The work by Mark Blyth (2011) will help us to understand the importance of ideas in human action, social systems and the polity. From Blyth stand point, it is not a collection of data, as accurate as it could be, or software enable technologies that would preclude future societies. The arguments run as follows. Mark Blyth (2011) explains that we might use ideas as a central point of explanations from several points of view, an empirical, a taxonomic and an additional one, where ideas are viewed as fundamental to both the nature of human action and causation in social systems. Blyth (2011) contends that explaining how the social world is put together necessitates a deep and systematic engagement with ideas. This is because without ideas neither stability nor change in social systems can be fully understood, and specially change. Blyth (2011) focuses on taken-for-granted assumptions that make non-ideational theories work, and should not automatically be taken for granted. Keynes has also suggested that the world of human action might be more ontologically different than the case appears at first blush, as Blyth explains: While those parts of the physical world constituted by observable fixed value generators and constant causes might be predictable to a large degree, "the social world more generally might be characterized by uncertainty rather than probabilistic risk (...). In such a world, past events and strategies drawn from them might not, (...) be a good guide to the future (Blyth, 2011, p. 88)."

Blyth (2011, p. 89), argues as follows: "Admitting that the world is deeply uncertain, rather than risky, is, however, problematic for any non ideational social science." In fact, when we accept uncertainty rather than risk, finite variance in outcomes cannot be assumed. This means consequently, that causation, parameter estimates, the central limit theorem, probability calculus, ordinary least squares, and even linear efficiency are all called into question. And if we live in a world of uncertainty rather than risk, ideas matter and become fundamental to any social science (Blyth 2011, p. 89). Going further and following the same argument, we find Taleb and Pilpel (2003) arguing about complex generators, or complex systems: "it is not that it takes time for the experimental moments . . . to converge to the 'true' [moments]. In this case, these moments simply do not exist. This means . . . that no amount of observation whatsoever will give us E(Xn) [expected mean], Var(Xn) [expected variance], or higher-level moments that are close to the 'true' values . . . since no true values exist" (Taleb and Pilpel 2003, 14). The view that Blyth proposes instead is "a world of inconstant and emergent causes, rather than a world of linear causation. In such a world, outcomes are truly uncertain rather than risky, since the causes of phenomena in one period are not the same causes in a later period (Blyth, 2011, p. 92)."

In complex systems, emergence and interdependence are the dynamics, while simple causal linearity is the exception (Blyth 2011 p. 93). Moreover, subject and object are not independent, but interdependent, because actions taken in light of beliefs alter the nature of the system itself (Cartwright 2007). In human action we find uncertainty, nonnormality, interdependence, and nonlinearity as equi-plausible conditions, and this is why ideas need to be central to social scientific endeavors (Blyth 2011, p. 94). We could ask ourselves whether uncertainty, nonnormality and nonlinearity could be accepted in artificial intelligence as sources of intended social changes, and the answer would certainly be it could not. Here we find a frontier, where only human value based judgments would be entitled to venture.

Taking the view of lawyers such as Ronald Coase (1974), and more so for historical institutionalists, paying attention to ideas, **ideas are much more than filters or fillers as we could contend that artificial intelligences are**. They are variously norms, conventions, schemas, and ideologies, collective products that make the world hang together (see Schmidt, 2011; Berman, 2011; and Lieberman, 2011, chapters 2, 5, and 10 in this book). According to Blyth they implicitly see the world as being more uncertain than risky (Blyth, 2011, p. 95). Following with these lines of thought, if the world is more uncertain, more nonlinear, and less normal than is commonly assumed, then a helpful way to view the environment is not to view it as a set of constant causes and invariant rules like a mechanical system, --that would be closer to an artificial intelligence, including software-enabled technologies and big data. Rather, Blyth (2011, p.96) suggests placing ideas in an uncertain world, in an evolutionary perspective, in which ideas, agents, and an uncertain environment codetermine one another (Blyth, p. 96). Social systems are most definitely complex adaptive systems replete with feedback loops, unintended consequences, and nonlinear dynamics. They are also decidedly unergodic and nonprogrammable (Kauffman 2008).

How we think about the world affects the strategies for building contingent stability in that world (Blyth, p. 98), and this is a strong reason to defend that software-enabled technologies and urban big data, while being expert tools cannot replace polities and decisions to conform social pacts and social arrangements. While software-enabled technologies and urban big data overdetermine what currently works, the future is in fact "underdetermining" (Blyth, 2011, p. 99). Ideas and environment can combine to produce outcomes that no one expects (Blyth, 2011, p 99).

Other reasons to defend the importance of human ideas revolve around interdependence and evolution. Blyth defends that the interdependence of subject and object in social systems is an absolutely fundamental aspect of their existence: "Not addressing this and instead speaking of idealized "independent" and "dependent" variables (subjects and objects) ignore their mutual constitution and imbrications over time and the evolutionary nature of social systems" (Blyth, 2011, p 101). Lewis and Steinmo also explain it: "the idea of isolating factors as independent variables may be an ontological fallacy" (2007, 10) that relies on reducing the world to a game of dice. Blyth goes on explaining:

> *"Unfortunately, from the point of view of predictive social theory, at least, we probably do not live in this world. Unlike ideas such as the "laws" of probability that concern fixed and isolated known-value generators such as dice, ideas about the workings of the social world are almost never correspondence theories of the world as it really is, since that world is always evolving. In the language of* ***philosophy, it is always becoming, never being.*** *The struggle to define the world and thereby delineating what is worth bothering about in the first place—inflation or unemployment, Iran or China, American Idol or Americans being idle—is a political struggle that is fundamentally a contest over ideas" (Blyth, 2011 p. 101).*

In the same vein, Helbing explains that having more information than humans (as cognitive computers have today) does not mean to be objective or right (Helbing, 2019). Moreover, big data approaches and the learning of facts from the past are usually bad at predicting fundamental shifts as they occur at societal tipping points (Helbing, 2019), and this is what social sciences mainly need to care about. Helbing defends that big data analytics often results in meaningless patterns and spurious correlations, for the sake of objectivity and in order to come to reliable conclusions, it would be good to view its results as hypotheses and to verify or falsify them with different approaches afterwards (Helbing, 2019, 15). And here, Helbing suggests some proposals for accountability. First of all, to ensure that scientific standards are applied to the use of Big Data: "For example, one should require the same level of significance that is demanded in statistics and for the approval of medical drugs. The reproducibility of results of big data analytics must be demanded." (Helbing, 2019, 15). Secondly, a sufficient level of transparency and/or independent quality control is needed to ensure that quality standards are met. Thirly, it must be guaranteed that applicable antidiscrimination laws are not implicitly undermined and violated. Fourthly, it must be possible to challenge and check the results of big data analytics. And lastly, efficient procedures to compensate individuals and companies for improper results, or in the case of data use particularly, for unjustified disadvantages (Helbing, 2019, 15). Helbing points at trust as an additional trait of accountability. For democratic governments, public trust would be the basis of legitimacy and power. For companies, the trust of consumers and users would be important to gain and maintain a large customer base (Helbing, 2019, 16). Trust, however, is a fuzzy concept if we are not able to link it to institutionalized forms of accountability.

VI. A fourth independent branch of government?

There is a big need to work on mechanisms to maintain, strengthen AI governance in democracies. As Harari states, we have a long experience regulating property and land, but we are at our infancy regulating artificial intelligence and data (Harari, 2014). Accountability is the aspect this work focuses more on, and in this section the key are aspects related to accountability in the first place. Secondly, the idea to think of a possible fourth independent branch of government is advanced.

In order to further accountability, Helbing et al. (2019) defend that basic rights of citizens should be protected. Basic rights are “a fundamental prerequisite of a modern functional, democratic society” (Helbing et al, 2019, 73). Following Helbing et al., this form of accountability would require the creation of a new social contract, based on trust and cooperation, with citizens and customers as partners, not as obstacles or resources to be exploited. For this, an appropriate regulatory framework shall be developed, which ensures that technologies are designed and used in ways compatible with democracy. Helbing et al. (2019) defend there should be a right in a simple way, to get a copy of personal data collected about us. Helbing et al. (2019) also propose that law regulates that this information must be automatically sent, in a standardized format, to a personal data store, through which individuals could manage the use of their data (potentially supported by particular AI-based digital assistants).

Insisting on accountability, to ensure greater privacy and to prevent discrimination, the unauthorised use of data would have to be punishable by law. In this way individuals would be able to decide who can use their information, for what purpose and for how long -including this time frame as compulsory could be an interesting way to ensure accountability. Appropriate measures should be taken too to ensure that data is securely stored and exchanged (Helbing et al, 2019) Interesting for an informed and educated citizen, Helbing et al. (2019) propose sophisticated reputation systems considering multiple criteria that could help to increase the quality of information on which citizen decisions are based. If data filters and recommendation and search algorithms would be selectable and configurable by the user, Helbing et al. (2019) suggest, we could look at problems from multiple perspectives, and we would be less prone to manipulation by distorted information (Helbing et al, 2019). In addition, there would be a need for an efficient complaints procedure for citizens, as well as effective sanctions for violations of the rules. Finally, in order to create sufficient transparency and trust, leading scientific institutions should act as trustees of the data and algorithms that currently evade democratic control, Helbing et al. (2019) propose. This would also require an appropriate code of conduct that, at the very least, would have to be followed by anyone with access to sensitive data and algorithms—a kind of Hippocratic Oath for information technology professionals (Helbing et al, 2019).

Furthermore, a digital agenda to lay the foundation for new jobs and the future of the digital society would be essential. As every year we invest billions in the agricultural sector and public infrastructure, schools and universities—to the benefit of industry and the service sector, the digital agenda shall be fundamental (Helbing et al, 2019). According to Helbing et al. (2019) the new public systems to ensure that the digital society becomes a success had to be based on new educational concepts, more focused on critical thinking, creativity, inventiveness and entrepreneurship than on creating standardised workers (whose tasks, in the future, will be done by robots and computer algorithms). Education should also provide an understanding of the responsible and critical use of digital technologies (Gil et al., 2015, Helbing et al, 2019).

A legal framework for automated technologies and intelligent machines is also a need: where autonomy needs to come with responsibility (Helbing, 2019, p. 17). Helbing et al. (2019) defend that humans should judge recommendations of super-intelligent machines, and put their suggestions in a historical, cultural, social, economic and ethical perspective. This could be done with students from the school, so that part of education includes algorithm literacy for a smart and responsible government. Because a smart government is not that in which citizens and educated groups in society refrain from governance. A smart government is that in which machines and artificial intelligences are understood and supervised by human institutions. In this vein, super-intelligent machines should be accessible not only to governing political parties, but also to the opposition (and their respectively commissioned experts), because the discussion about the choice of the goal function and the implication of this choice is inevitable.

This is where politics enters in times of evidence-or science-based decision making. And we could even think of a fourth independent government branch, apart from the executive, the legislative and the courts, with the endowment to ensure that algorithms in today's´democracies are responsible to the principles of Constitutions at different government levels, from cities, to regions to national and supranational government constructions. A precedent for a regulatory framework in which different levels of government participate in the regulation of the market exists for instance in the field of telecommunications regulation in the United States, where even the local level, through Public Utilities Commissions (Gil, 2002, 15) and a national association of commissioners exists from 1889 (National Association of Regulatory Utility Commissioners, NARUC), and there is a specific committee in the Federal Communications Commission , the FCC Local Government Advisory Committee balancing the relations among the FCC, the local and federated governments (Gil, 2002, 21).

In the composition of this independent government branch with the endowment to ensure that algorithms in today's´democracies are responsible to the principles of Constitutions, there could be from data scientists to philosophers, lawyers and common citizens, and it would have the capacity of furthering accountability of AI and make proposals for AI literacy at all levels of society. This proposal would go further than the existing and ongoing relations of NARUC and NARUC affiliates globally, in which experts including lawyers and engineers participate, to devise mechanisms including broader citizen participation. If we through light over current affiliates of NARUC,[2] beside from the five NARUC U.S. regulatory affiliates,[3] there are international organizations, national regulatory associations, regional conferences,

[2] National Association of Regulatory Utility Commissioners (NARUC) https://www.naruc.org/regionals/

[3] There are states organizations that are either part of NARUC or have institutional relations: the Organization of MISO States, Inc. (OMS), the Organization of PJM States, Inc. (OPSI), the Southeastern Association of Regulatory Utility Commissioners (SEARUC), and the Western Conference of Public Service Commissioners (Western). Further mechanisms allow for coordination, such as The Mid-America Regulatory Conference (MARC), the Mid-Atlantic Conference of Regulatory Utilities Commissions (MACRUC), the National Association of Pipeline Safety Representatives (NAPSR), the National Conference of Regulatory Attorneys, the National Conference of Regulatory Utility Engineers, the National Conference of State Transportation Specialists (NCSTS), the National Regulatory Research Institute (NRRI), and the New England Conference of Public Utilities Commissioners, Inc. (NECPUC).

and regional states organizations. Starting with the two international organizations as well as smaller, informal bodies that focus on specific industry issues worldwide, and have formal relations with NARUC, we find Canada's Energy and Utility Regulators (CAMPUT), the East Asia and Pacific Infrastructure Regulatory Forum (EAPIRF), the Energy Regulators Regional Association (ERRA), the International Confederation of Energy Regulators (ICER).

VII. Conclusions and further research

In this work, insights from different perspectives, from philosophy of science to political theory, public administration, political science, institutional law and data scientists have helped us to tackle the issue of accountability in artificial intelligence development. Bringing together the insights from above specialists streams of science the work shows that:

- Artificial intelligence is a contested contribution to governance per se. Many specialists prefer to envisage AI as an aid for governance, where value based judgements are key in making decisions both in the public and the private sectors. We have also seen that we are at a very early stage in the governance of AI. Classical control engineering, resolved within firms (Müller, 2019), which are hierarchies, cannot be translated to societies that are democratic political economies, where decisions are legitimized horizontally -through votes.
- The widespread belief that large datasets offer a higher form of intelligence and knowledge, with the aura of truth, objectivity and accuracy is a myth. This particularly is a claim of data scientists relating artificial intelligence and big data to societal problem solving from a critical perspective. This is also linked to claims from philosophical perspectives that truth cannot be guaranteed through method (Ongaro, 2019). Furthermore, the need for a critical perspective is also linked to uncertainty -instead of probability- as a feature of the social world, where past events and strategies drawn from them might not be a good guide to future scenarios (Blyth, 2011). More generally, it would contend with the philosophical idea that the world is always becoming, never being.
- Artificial intelligence is a contested, evolving and unresolved concept. As definitions matter, some definitions argue that we should include within robots humanoid companions, dataveillance, and algorithms. Robotic devices will be more common outside industry, in any smart system, from the home to hospitals and public places in the city. This could bring us to a re-definition of what a fourth independent government branch shall include, perhaps all of them: humanoid companions, dataveillance, and algorithms.
- Artificial intelligence cannot subdue the need to rethink government priorities. Instead, AI is a good opportunity to revisit priorities at the local, national, and supranational levels. This is linked to the work section on human ideas about the future, ideas that have always been key to building up the future of societies. Human ideas and beliefs alter the nature of the system itself (Cartwritght, 2007), while in human action we find uncertainty, nonnormality, interdepence and nonlinearity (Blyth,

2007), traits that we would not want to code in an artificial intelligence. Ideas, the work defends, are much more than filters or fillers as we could contend that artificial intelligence is.

- Departing from morality, the work defends the position of authors advancing a need of further discussion on practical accountability, especially focusing on the network of individuals programming, and on aspects attached to legal and policy conflicts (Stamboliev, 2019). Helbing et al. (2019) have proposed an appropriate code of conduct for anyone with access to sensitive data and algorithms—a kind of Hippocratic Oath for information technology professionals (Helbing et al, 2019) and efficient complaints procedure for citizens, as well as effective sanctions for violations of the rules. Going further, Helbing et al. (2019) put forward proposals such as 1) demanding the same level of significance in statistics than for the approval of medical drugs; 2) the reproducibility of results of big data analytics 3) a sufficient level of transparency and/or independent quality control, needed to ensure that quality standards are met; 4) guarantees that applicable antidiscrimination laws are not implicitly undermined and violated; 5) it must be possible to challenge and check the results of big data analytics; 6) efficient procedures to compensate individuals and companies for improper results, or in the case of data use particularly, for unjustified disadvantages. Helbing et al. (2019) also defend trust as an additional trait of accountability.
- As humans should judge recommendations of super-intelligent machines, and put their suggestions in a historical, cultural, social, economic and ethical perspective, there is a need to have AI literate citizens, to avoid further digital divides. And this could be done with students from the school, so that part of education includes algorithm literacy for a smart and responsible government, and with specific programs for minorities and special groups in society.
- The work finally defends that judging the recommendations of superintelligent machines could also be done through the articulation of citizen participation in a fourth independent branch of government, in charge of ensuring that algorithms in today's´democracies are responsible to the principles of Constitutions at different government levels, from cities, to regions to national and supranational government constructions. In the composition of this independent government branch there could also be data scientists, philosophers, lawyers as well as common citizens. This fourth branch would have the capacity of furthering accountability of AI and make proposals for AI literacy at all levels of society.

References

BEKMAMEDOVA N., SHANKS G. Social media analytics and business value: a theoretical framework and case study. In: Proceedings of 2014 47th Hawaii international conference on system sciences (HICSS). IEEE. 2014.

BEYER, MA, LANEY, D. The importance of ‘big data’: a definition. Gartner, Stamford. 2012. p. 2014–2018.

BHARADWAJ, A., EL SAWY, O, PAVLOU, P. A. AND VENKATRAMAN, N. "Digital business strategy: toward a next generation of insights." *MIS quarterly.* 2013, p. 471-482.

BLYTH, M. Ideas, uncertainty, and evolution. *Ideas and politics in social science research*. Oxford University Press, 2011, p. 83-101.

BOYD, D., and CRAWFORD, K. "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon." *Information, communication & society* 15, no. 5, 2012, 662-679.

CARTWRIGHT, N. *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge University Press, 2007.

CATH, C., WACHTER, S., MITTELSTADT, B., TADDEO, M., & FLORIDI, L. . Artificial intelligence and the ‘good society’: the US, EU, and UK approach. *Science and engineering ethics*, 24(2), 2018, p. 505-528.

COASE, R. H. The market for goods and the market for ideas. *The American Economic Review*, 64(2), 1974, p. 384-391.

CONSTANTIOU, I. D., and KALLINIKOS, J. "New games, new rules: big data and the changing context of strategy." Journal of Info*rmation Technology* 30, no. 1, 2015, p. 44-57.

CORTÉS-CEDIEL, María E.; GIL, Olga; CANTADOR, Iván. Defining the engagement life cycle in e-participation. Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age. 2018. p. 1-2.

DAVIS, C.K. Beyond data and analysis. *Communications of the ACM.* 57(6), 2014, 39–41

European Parliament Committee on Legal Affairs. Civil law rules on robotics (2015/2103 (INL). Brussels, Belgium: European Parliament. 2016. Retrieved on December 13th, 2019 from: https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf

European Union. European Union (EU) General Data Protection Regulation 2016/679. Brussels, Belgium. 2016. Retrieved on December 13th, 2019 from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Executive Office of the President. Artificial intelligence, automation and the economy. Washington, DC, USA. 2016. Retrieved on December 13th, 2019 from: https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

Executive Office of the President National Science and Technology Council Committee on Technology. Preparing for the future of artificial intelligence. Washington, DC, USA. 2016. Retrieved on December 13th, 2019 from: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

FLORIDI, L., & TADDEO, M. What is data ethics? *Philosophical Transactions of the Royal Society.* 2016. https://doi.org/10.1098/rsta.2016.0360

GALLIE, W. B. Essentially contested concepts. Proceedings of the Aristotelian society. 56, 1956, 167–198. doi:10. 1093/aristotelian/56.1.167

GANTZ, J, & REINSEL, D. The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future.* Vol. 2007. 2012, p. 1–16

GIL, O., NAVÍO, J., & PÉREZ DE HEREDIA, M. *¿Cómo se gobiernan las ciudades? Ciudades inteligentes. Casos comparados: Shangái, Iskandar, ciudades en Japón, Nueva York, Ámsterdam, Málaga, Santander y Tarragona.* 2015. Tarragona: Silva Editorial.

GIL, O. Telecomunicaciones y política en Estados Unidos y España (1875-2002): construyendo mercados. Madrid, Siglo XXI de España Editores, 2002.

GLASMEIER, A., & NEBIOLO, M. Thinking about smart cities: The travels of a policy idea that promises a great deal, but so far has delivered modest results. *Sustainability*, *8*(11), 2016, p. 1122.

HELBING, D., FREY, B. S., GIGERENZER, G., HAFEN, E., HAGNER, M., HOFSTETTER, Y., & ZWITTER, A. Will democracy survive big data and artificial intelligence? *Towards Digital Enlightenment.* 2019, pp. 73-98. Springer, Cham.

HELBING, D. Societal, economic, ethical and legal challenges of the digital revolution: from big data to deep learning, artificial intelligence, and manipulative technologies. En *Towards Digital Enlightenment*. Springer, Cham, 2019. p. 47-72.

HILL, R. K. What an algorithm is. *Philosophy & Technology,* 29(1), 2016, pp. 35-59.

KITCHIN, R. The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society.* 2016, 374: 20160115. http://dx.doi.org/10.1098/rsta.2016.0115

MATHESON, C. Weber and the Classification of Forms of Legitimacy. *British Journal of Sociology*, 1987, 199-215.

MIKALEF, P., PAPPAS, I. O., KROGSTIE, J., & GIANNAKOS, M. Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management*, 16(3), 2018, 547-578.

MULLIGAN, D.K., KOOPMAN, C., DOTY, N. Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. *Philosophical Transactions of the Royal Society*. 2016. A374: 20160118. http://dx.doi.org/10.1098/rsta.2016.0118

MÜLLER, V. C. Ethics of Artificial Intelligence and Robotics.Stanford Encyclopedia of Philosophy (Palo Alto: CSLI, Stanford University. Forthcoming. http://plato.stanford.edu/ Las retrieved 17/01/2020 from https://philarchive.org/archive/MLLEOA-4

NYHOLM, S. Attributing agency to automated systems: reflections on human–robot collaborations and responsibility-loci. *Science and engineering ethics*, 24(4), 2018, pp. 1201-1219.

RUSSOM, P. Big data analytics. *TDWI Best Practices Report*, Fourth Quarter, 2011, pp. 1–35

SCHROECK M, SHOCKLEY, R., SMART, J., ROMERO-MORALES D., TUFANO, P. Analytics: The real-world use of big data. IBM Global Business Services, 2012, pp. 1–20

SHAHROKNI, H., & SOLACOLU, A. Real-time ethics—A technology enabled paradigm of everyday ethics in smart cities: shifting sustainability responsibilities through citizen empowerment. In *2015 IEEE International Symposium on Technology and Society (ISTAS)*, IEEE, 2015, November, pp. 1-5.

SHARIFI, A. A critical review of selected smart city assessment tools and indicator sets. *Journal of Cleaner Production.* 2019.

STAMBOLIEV, E. Challenging Robot Morality: An Ethical Debate on Humanoid Companions, Dataveillance, and Algorithms (Doctoral dissertation, University of Plymouth). 2019.

SUN, E.W., CHEN, Y.T., YU, M.T. Generalized optimal wavelet decomposing algorithm for big financial data. *International Journal of Production Economics*. 165, 2015, pp. 194–214

WHITE, C. Using big data for smarter decision making IBM. Yorktown Heights, New York, 2011.

YIGITCANLAR, T, et al. "Can cities become smart without being sustainable? A systematic review of the literature," *Sustainable cities and society*, 2018.

OTHER INTERNET SOURCES

Canada's Energy and Utility Regulators (CAMPUT) http://www.camput.org/

East Asia and Pacific Infrastructure Regulatory Forum (EAPIRF) http://www.eapirf.org/

Energy Regulators Regional Association (ERRA) http://www.erranet.org/

International Confederation of Energy Regulators (ICER) http://icer-regulators.net/

Mid-America Regulatory Conference (MARC) http://www.marc-conference.org/

Mid-Atlantic Conference of Regulatory Utilities Commissions (MACRUC) http://macruc.org/

National Association of Regulatory Utility Commissioners (NARUC) https://www.naruc.org

National Association of Pipeline Safety Representatives (NAPSR) http://www.napsr.org/

National Conference of Regulatory Attorneys https://www.naruc.org/about-naruc/event-calendar/2020-national-conference-of-regulatory-attorneys/

National Conference of Regulatory Utility Engineers

National Conference of State Transportation Specialists (NCSTS) http://ncsts.naruc.org/

National Regulatory Research Institute (NRRI) https://www.naruc.org/nrri/

New England Conference of Public Utilities Commissioners, Inc. (NECPUC) http://necpuc.org/

Organization of MISO States, Inc. (OMS) https://www.misostates.org/

Organization of PJM States, Inc. (OPSI) https://opsi.us/

Southeastern Association of Regulatory Utility Commissioners (SEARUC) http://www.searuc.org/

Western Conference of Public Service Commissioners (Western) http://western.naruc.org/