Data-driven democracy: Who decides?

Margarita Gómez-Reino Cachafeiro
Departamento de Ciencia Política y de la Administración
UNED
Madrid, Spain

Alberto Suárez
Machine Learning Group
Computer Science Department
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Madrid, Spain

Abstract: Political regimes have evolved since Aristotelian times and we face today the global dominance of democratic regimes. One of the pillars of contemporary democracies is the right to participate in elections and to choose representatives. During the XXth century censitarian systems gave way to universal franchise, which included women's right to vote. In this manner, the participation in the political decision process through voting became one fundamental mechanisms for democracy, as it is now known. However, which kind of politics can we expect in a world dominated by Artificial Intelligence? What types of democracies will emerge? In the short story 'Franchise' written by Isaac Asimov in 1955, machines are crucial for the definition of collective choices. In the society envisioned by Asimov, universal suffrage is limited to one single citizen chosen by a computer as representative of the political community. However, the final decision is made by the computer on the basis of large amounts of data. The contribution of the citizen is but another input in this algorithmic decision process. As Harari argues democratic principles are threatened when democracy is defined as a data processing system. In this work we provide a review of current AI systems, especially those that are data-driven, to assess the plausibility of Asimov's scenario. We conclude that current AI systems lack the generality that such a comprehensive computerized decision system would require. We finally provide a critical analysis of how AI systems can be incorporated in the political decision process.

Keywords: data-driven democracy, Artificial Intelligence, Franchise, decision processes

1. Introduction

In 1955 Isaac Asimov published a short science fiction story entitled 'Franchise'. In the story, set in the USA, a dystopian presidential election process is described. A single voter, selected by computers, can by himself elect the Presidential winner thanks to the assistance of the powerful Multivac computer. The story describes the transition from elections in which everybody exercised the right to vote (in liberal democracies), to a new system in which a single representative citizen, with the assistance of a computer, actually votes. The mystery of the election under one single voter is an anti-climatic process with two main actors, a citizen from Indiana, named Muller, and a special machine, Multivac. The machine already has most information but it needs to ask the protagonist a few questions about his attitudes and feelings to reach a final decision. One individual representing all, as the story concludes that 'the sovereign citizens of the first and greatest Electronic democracy had through Norman Muller exercised its free and untrammeled franchise' (Asimov 1955:15). This is a highly distressing conclusion, since Muller does not formally cast his vote, or has the opportunity to express his opinions or values through the election. He is a passive agent that provides but a few among a myriad of inputs to a computer so that a decision can be reached. Asimov's future (2008 in the fiction), has already come into existence and so have some features of this imaginary data-driven democracy. Nevertheless, the current expectations about the role of Artificial Intelligence in democratic systems, and in particular, in decision making in democracies, are possibly too high.

This fiction is the starting point for our discussion. In this paper we explore the compatibility of the principles of AI and democratic systems in decision making processes. The structure of the paper is as follows: First, we introduce the principles of democracy and democratic decision making under majority rule. Second, we define and describe the AI and its current state of the art. We conclude the discussion with a critical discussion on the role of current AI systems for decision making processes.

2. Democracies, voting and decision making process

Political Science has paid special attention to the definition and concept of political regimes. Since Aristotle presented his typology of virtuous and vicious regimes, constructed upon two criteria: who governs and how this government is exercised, one of the central pillars of academic work has been devoted to the definition of political structures, in particular the historical development of liberal (representative) democracies of the Western type.

Table 1. Aristotle's typology of regimes

	The one	The few	The many
Common interest	Monarchy	Aristocracy	Polity (Republic)
Rulers' self-interest	Tyranny	Oligarchy	Democracy

Until the 1960s the literature commonly divided regimes as democracies and totalitarian states. In 'Problems of Democratic Transition and Consolidation' Linz and Stepan (1996) elaborated pre-existing typologies, to further add two non democratic types within the context of the so called third wave of democratization (Huntington 1992). The expectation was that democracy became the dominant political regime in the world.

Central to the principle of democratic systems is the idea that a majority rules and individual values and tastes are determinant to the construction of a social welfare function by aggregating individual choices (Arrow 1963:103). According to Arrow, in a capitalist democracy there are two methods by which social choices can be made: voting to make political decisions and the market mechanism to make economic decisions (Arrow 1963:1). Borda, Condorcet and Arrow, among others, have theorized on the methods to arrive collective decision making and the problems and paradoxes from individual preferences to collective ones (Arrow 1963).

Harari argues that political structures are increasingly defined as systems for data processing (Harari 2017:406). In his typology of political regimes, he asserts that the main difference between dictatorship and democracy is the centralized or distributed data processing systems, the latter performance more functioning (p.406). Harari joins Asimov in his future expectations about the basic erosion of democratic systems. As the conditions for data processing will change in the XXI century, democracy could decay and even disappear as volume and speed grow, institutions such as elections, political parties and parliaments would become obsoletes' (Harari 2017:406). Harari's point is not based on ethics, rather on the efficiency of alternative methods. Technological revolutions have surpassed political processes, both parliaments and voters lose political control. In Harari's view, within the next decades more revolutions, similar to internet, will take place in which technology will win over politics. For Harari 'traditional democratic politics loses control over events as it cannot provide meaningful visions of society' (Harari 2017:408).Thus, we face a situation of power vacuum in which cutting edge technology goes hand in hand with myopic politics. If traditional political structures cannot process data, he argues, then new political structures will appear out of evolution and without a resemblance of dictatorship and democracy.

Asimov's story puzzles us, eliciting a reflection on the shaky boundaries between dictatorship and democracy. Democracy is a political regime rule by the people through free and fair elections. The fundamental of democracy is popular sovereignty, the idea that people are the ultimate authority and the source of authority. Free elections are thus essential to democracy. Franchise is a right to vote in the election of public officials, not a computerized decision process based on of probing the average citizen's attitudes and feelings, as in Asimov's tale.

Democracy since the French Revolution is associated with majority rule (and minority rights). Individual values are here taken as data and are not susceptible of being altered by the nature of the decision process itself (Arrow 1963:7). Arrow considers that the process by which society makes its choice is a value in itself (Arrow 1963: 89). According to Arrow, there are different mathematical forms of the social utility function in terms of individual utilities or their product or the product of their logarithms, or the sum of their product, taken two at a time (4).

In the following section, we provide an overview of AI systems in order to provide a basis of discussion for their integration in political decision processes.

3. Decision systems in Artificial intelligence

In its current embodiment, the (AI) project is identified with the design and implementation of computational systems exhibiting behavior that can be thought of as intelligent. To avoid the quasi-tautological character of this definition and a lengthy, possibly barren, discussion on what exactly constitutes intelligence, Alan Turing proposed to use an operational test (Turing, 1950) in which a human being acts as a judge to determine whether the implemented system exhibits intelligent behavior. In this *Imitation Game*, the name originally used to denote what is currently known as the Turing test, a human, whom we assume to be an intelligent agent, interacts with another human (i.e., an intelligent agent as well) and with a computational system. The interaction takes place in a manner such that characteristics that are deemed irrelevant for intelligence (e.g. the appearance of the computational system, its ability to speak with a human inflection, etc.) do not interfere with the assessment. In its most common form, this interaction takes place through a computer chat in which the three agents involved exchange text messages. If, after a reasonable amount of time, the human is unable to distinguish between the other human and the computational system, one concludes that the machine has passed the Turing test, and can therefore be said to exhibit intelligence.

This operational definition of intelligence has been amply criticized on different grounds. Most of these critiques hinge on the observation that, while it may well be possible to build a machine that *imitates* intelligence, and therefore passes the Turing test, such a machine cannot be said to *be* intelligent. Irrespective of the fact that a strict essentialist position on intelligence is difficult to uphold without falling into some kind of solipsism (in particular, one could also apply the objection to humans, or at least of other humans, arguing that they are in fact not intelligent, that that thet simply imitate intelligence), the question of whether imitation is sufficient proof of intelligence deserves some attention with regard to the question addressed in this paper. In particular, Searle (1980) argues that *intentionality*, which he assumes to be a result of the causal functioning of the brain and a defining characteristic of biological intelligence, cannot be the result of the instantiation of a program in a computer. According to this argument the strong AI project cannot be realized simply by designing programs. Instead, one would have to replicate the causal powers of the brain.

Analyzing the empirical evidence on the level of current progress attained, it is clear that we are still far from the point at which one can convincingly claim that an artificial system has passed the Turing test. In spite of the impressive performance of some AI systems, the project is still in its infancy, far from the level of generality that is a necessary condition for the *strong AI* project. At the moment, only weak forms of AI are possible. The computational systems developed exhibit proficiency at specific tasks (e.g. play chess, translate texts, or describe scenes in image), but are utterly inept at other tasks, even some that are closely related (e.g. play checkers, classify texts, or label faces in a picture). For instance, it is relatively simple to build a system, say, e.g., a decision tree or a neural network, which, on the basis of labelled examples (i.e., the system is trained on images of triangles and hexagons that are identified as such), learns to discriminate between hexagons and triangles. However, if we present the system with a new shape, say a circle, the system will not be able to tell that this shape deviates from the previous ones, unless we specifically program that capacity. By contrast, in many such situations, humans are able detect the novelty in the presented shape, find potential uses, even generate a different label to refer to this new concept, and incorporate it into her world view. These capacities of receptiveness and adaptation to change, generation of meaning, and integration of into a comprehensive knowledge system are currently lacking in our highly specialized AI systems. One can say that the systems developed are but syntactic constructs (programs that encode specific algorithmic procedures) that are not sufficient to represent semantics.

In spite of its recent identification with Computational Intelligence, the AI project is grounded on antecedents that predate the modern computer. In our quest for self-knowledge, most of us would recognize intelligence as one of the defining elements of the human condition. Because of the similarities of some behavioral patterns with humans, we do ascribe some degree of intelligence to other animals, especially apes and, to a certain extent, mammals. We also metaphorically speak of ant and bee colonies as having some form of collective intelligence. However, humans often reject the idea of intelligence in inanimate objects and machines. In some cases, the project of developing an artificial system endowed with intelligence is met with incredulity, disgust, or even disapproval on moral grounds. It is not uncommon to hear expressions such as 'A computer will never be able to X’, where X stands for activities that humans characterize as intelligent. Instances of X are: 'play chess’, 'drive a car’, 'diagnose a disease', 'translate texts', 'prove a mathematical theorem', 'tell a joke', 'express feelings', 'write a poem', etcetera. From these tasks, computational systems have been developed that exhibit a definite mastery at the first two. Computers can achieve human-level performance at some problems in medical diagnosis, lingüistics and formal reasoning. The level of success in the last three tasks is more difficult to determine, because humor, emotions, and art need be assessed not only from the intention the agent that is the source of the message, but also from the response that such a message elicits in the receiver. Curiously enough, once a machine becomes more proficient than a human at a particular task, the solution of the task is viewed as merely requiring mechanical skills, other than intelligence: raw computational power, access to large amounts of information, and so on. The algorithmic nature of the solution is interpreted as being devoid of the quality and depth that we ascribe to human intelligence. A paradigmatic illustration is the decline of the intellectual prestige of chess. The reputation of chess achieved its peak in the Cold War period, barely survived Kasparov’s 1997 defeat at the hands of IBM’s Deep Blue, and has recently suffered a further blow at the hands of Google’s entirely self-taught alphaZero chess program (Silver et al 2018).

In spite of these reservations, from the efforts to mechanize thought in Aristotelian logic, to the design of mechanical control systems, clocks, mechanical calculators, and automata that imitate animal and human behavior, humankind has persistently endeavored to design artificial systems that exhibit intelligence to some extent. The origins of the modern AI project can be traced to the aftermath of World War II. Of particular importance are the theoretical contributions of Alan Turing, who in his 1950 article already identifies learning and adaptation as central aspects of the AI project. In that article Turing proposes to address design problems through a simulated evolution process, which eliminates the need of a prespecified plan or a designer. In such a simulation a population of artificial systems that exhibits some diversity (e.g. each individual in a population is a variant of some algorithmic procedure) is made to evolve by mechanisms that introduce variability in the population (e.g. changes in the algorithm akin to mutations, or exchanges of segments of code among individuals, similar to the exchange of DNA that takes place in the sexual reproduction of biological systems) in combination with a selection process based on the performance of the individual (e.g. a measure of the effectiveness of the solution provided by the corresponding program).

The unofficial birth certificate of the discipline is a research proposal by McCarthy, Minsky, Rochester, and Shannon (1955) for a 2 month, 10 man study of Artificial Intelligence to be carried out during the summer of 1956 at Dartmouth College. Starting from the conjecture that 'every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it' (McCarthy et al 1955), the goal is to explore how this conjecture can be realized in practice. In particular, the authors address questions such as how can a computational system be programmed to form concepts and abstractions, use a language, improve its capacities, or use controlled randomness to emulate creativity.

The early AI projects addressed stylized problems in small, controlled environments with little or no uncertainty. An example of the types of systems developed in this period is the *General Problem Solver* developed by Newell, Shaw, and Simon (1959) for the solution of a restricted class of problems in formal logic. Even though the system was in principle able to solve any problem of the type considered, in practice, the solutions to real-world problems were not accessible. The difficulty resides in the exponential increase of the time needed to make the necessary computations as the problems grow in size and complexity. Even if these computations were possible, the solution of a problem by means of logical inference requires a formalization both of the question to be addressed and of the knowledge base that is relevant for the solution. Typically, this formalization is made by a human expert, who needs to define an ontology for the resolution of the problem. The expert identifies the objects and relations that are relevant for the solution of the problem at hand. These are then encoded as symbols that allow us to build atomic statements with a definite truth value (e.g. 'A triangle is defined by three points in space'). These elementary statements can be articulated with the help of logical connectives (e.g. 'not', 'and', 'or', 'if… then...', 'if and only if...then…') to compose more complex statements (e.g. 'If three points form a triangle, the sum of angles defined by the segments that link them in pairs add up to 180º'). Once this has been done, deduction is a purely mechanical manipulation of symbols according to fixed inference rules. The process does not require any specific knowledge of the domain in which reasoning takes place. The purely syntactic nature of logical deduction, which renders its mechanization possible, implies that the solution to the problem posed is already contained in an implicit form in the formalized knowledge base. Since it is the human expert who formulates the problem in the formal language of problem, and interprets the output as a meaningful answer to the problem, it does not seem to be reasonable to ascribe intelligence to the formal deduction system, no matter how sophisticated or powerful is.

As the discipline matured, AI researchers started to realize that the limitation of resources, the open nature of the world, and the presence of uncertainty are intrinsic features of most problems that need to be addressed in practice. In consequence, more complex models that take these limitations in our knowledge and computational capacities were designed. Specifically, large systems of rules, called expert systems, were developed in an attempt to replicate the decision taking process by human specialists. In these systems, formal reasoning is replaced by numerical computations based on heuristic rules. However, it is very difficult to formalize the decision process of a human expert. In their reasoning, humans manage to integrate large amounts of information from various sources that may be contaminated by noise, include errors, and exhibit uncertainty. We have also mechanisms to focus our attention on crucial pieces of evidence that can be vague or hidden. Furthermore, the decision process is generally an interactive process in which information is acquired gradually and unlocks new avenues of exploration. It is therefore difficult to educe a static system of rules that replicates the human decision process. In a limited range of applications such rule-based expert systems provide excellent support for decisions. However, they require careful design and the fine-tuning of a large number of free parameters. This means that they are costly to build. They have also the disadvantage of being highly specific for the task for which they have been crafted. Typically, these systems are not robust to noise, cannot be easily scaled to handle to large problems, and have limited capacity of adaptation and self-improvement. Most recent approaches to AI are based on machine learning. In machine learning, concepts and relations are extracted by automatic induction from data. Instead of adopting a deductive scheme, as in formal logic reasoning, or attempting to encode the knowledge of an expert as systems of rules, one seeks to identify statistical regularity patterns in the data that can be exploited for analysis, modelling, and prediction. For the sake of concreteness, we will focus on the problem of classification, which is a type of *supervised learning*. In classification problems, a collection of labelled examples, the training data, is available for learning. Each of these data instances consists of a description of the object to be classified together with its class label. The goal of classification is to formulate, on the basis of the available data, a hypothesis that allows one to predict the class label of an unseen instance, which is characterized by its description only (i.e., its class label is unknown). The capacity to make accurate predictions on new instances that are independent of the ones used for learning is referred to as the *generalization capacity* of the predictive system.

The design cycle of such a classification system consists in collecting labeled data in sufficient quantities so as to make reliable induction possible. Then, a general model that makes minimal assumptions on the structure of the solution is formulated. The model parameters are determined through a learning process that makes use of the training data. Typically, learning proceeds by minimization of a cost function. The cost function usually involves a loss term that quantifies the accuracy of the predictions of the model on the training data and a term that penalizes the complexity of the learned model. The role of this second *regularization* term is to prevent the system from learning an overly complex model based on spurious patterns in the training data (fluctuations or errors) that are not useful and, in some cases, can be detrimental for generalization.

Decision trees and neural networks, including those with a deep architecture, are examples of prediction systems built in this manner. A decision tree is a hierarchical questionnaire. The first question in the hierarchy is the one that yields the largest amount of information on the class label based on the values of the attributes that characterize the examples. One typically explores only the space of questions that involve a single attribute. In binary decision trees an example is assigned to the right or the left branch of the tree, depending on the answer to the first binary question in the hierarchy at the root of the tree; for instance, the example is assigned to the right branch if the answer is 'yes', to the left branch if the answer is 'no'. If the class labels of the training examples assigned to one of the (left or child) branches are the same, then one does not need to use further questioning to discriminate among classes. By contrast, if in the branch considered the classes are mixed, one would need to further partition the data using, at least, an additional question. Note that a sequence of responses to the questions along a particular branch of the tree characterizes a subset of examples from the training data. The partitioning process proceeds until the subsets of training examples identified by the sequence of questions in the hierarchy are sufficiently pure (i.e. their class labels coincide) or no more questions can be posed. Note that the segmentations made on the basis of some of the final questions in the hierarchy may not be very relevant to the classification problem at hand because they are made on the basis of only a few training instances. To avoid spurious effects arising from these questions, the complexity of the model is reduced by sequentially eliminating some of the questions that appear at the bottom of the hierarchy. Whether or not a terminal question is eliminated depends, on the statistical significance of the partitions it gives rise to. This process is referred to as cost-complexity pruning and reflects an inductive bias towards simplicity. The fact that the induces decision tree is a hierarchical questionnaire means that we have a symbolic representation of the knowledge extracted from the data. The fact that these models are interpretable makes very attractive, in spite of the fact that their accuracy is not as good as that of sub-symbolic methods, such as neural networks.

Neural networks are processing systems based on the joint action of a collection of artificial neurons disposed in a network with specific connections. Each single neuron in the network carries out a simple processing of the signals it receives from other neurons modulated by the synaptic weights and outputs a signal for ulterior processing. For the sake of concreteness, we will focus on a specific type of network, called the Multilayer Perceptron (MLP). The first layer of an MLP takes as inputs the descriptors of the example to be classified. The last layer outputs a class label or an estimate of the posterior probability of the example to be classified. The remaining neurons are organized in a layered structure. Learning takes place by adjusting the synaptic weights between neurons in adjacent layers. The final knowledge representation, encoded by the values of the of the weights that have been learned, is distributed and sub-symbolic. The reason for this is that the values of the weights are numerical. Furthermore, it is hard to provide an interpretation for the individual weights. In fact, this is one of the disadvantages of neural networks with respect to decision trees. While they are generally more accurate, and flexible (e.g. they can adapt to 'concept drift', a term that refers to the gradual evolution in the relation between the attributes that describe the examples and the target we want to predict), the yield opaque models that are difficult to interpret and provide little or insight into the prediction problem considered.

Even though the role of the human expert in the machine learning process is more limited that in the previous approaches to reasoning, it is also fairly important: One needs to determine both the attributes that are used to characterize the different instances and the class labels. The attributes should be relevant for the task at hand. The class labels need to be accurate as well.

Once the characteristics of current AI systems have been briefly introduced, in the next section we will examine the difficulties associated with their integration in the political decision process, with a particular focus on the ones that are data-based.

4. On the use of AI in political decision processes

It is apparent that the use of AI tools, now incipient, will be pervasive in all areas of society. Politics will not be immune to this trend. In the dystopian future presented in Asimov's 'Franchise' Multivac is a powerful computer that decides the results of elections using an unintelligible algorithm that processes large amounts of data. The only direct human input is an interview by the computer of a person who is chosen by some obscure process by the computer as representing the 'average' citizen. Contrary to the expectations of the selected citizen, who is initially overwhelmed by the responsibility he imagines is associated to his franchise, rather trivial information is collected in the interview. In response to the citizen's perplexity, he is told that the input from the human is woven into 'the trillions of items it [Multivac] has' to reach a final decision. One is left to think that the citizen is only intended as a way to provide legitimacy to such a dehumanized decision process.

In spite of the naïveté of the story, Asimov presents us with some important questions about the nature of democracy, how citizens are engaged in the decision process, and the role of voting in an electronic democracy. With regard to democracy, at least two aspects of the story deserve some detailed analysis. The first one is related to the nature of representation in democracy. One of the goals of the political system is to unify in some way the diverse views and objectives of different groups in society in a manner that is conducive to the common good, while respecting the rights of minorities and of the individuals. A mechanism for dealing with this diversity is to establish institutions, such as the Parliament, that reproduce the composition of the society it represents. The decisions taken by these representative institutions can then be achieved by consensus, if the wills of the different groups happen to agree, or by negotiation, in which groups make concessions to reach an agreement. An alternative mechanism relies on the normalization of differences. This normalization can be achieved either by reducing the differences themselfs (e.g. through the creation of citizens who share a common perspective on the societal project), or by ignoring groups or individuals that deviate from the norm (e.g. governing for the 'silent majority', the 'average citizen', or, in some cases, 'the people').

The second issue is how to strike a balance between humans and technology in the decision-making process. While high-level decisions regarding the values and goals of a society should have the input of the individual citizens, the translation of these decisions into actual policies is in general too complex and time-consuming for direct participation. While most would agree on the need to involve the citizens in a society as closely as possible in the decision-making process, it is unavoidable that, at some point, the technical intricacies are too complex for a public debate to be fruitful. The question here is how to avoid creating an opaque layer of expert, or technology-based decision-making that supplants the necessary public debate.

From the review of the state of the art of the field of Artificial intelligence given in the previous section, it is apparent, that, while in some specific tasks computational systems exhibit a performance that is comparable or better that humans, we are far from having built a comprehensive device, such as Franchise's Multivac, capable of taking autonomous decisions without dedicated human intervention even in relatively simple situations. Even with the reduced functionality of current AI systems, there are some important issues that need to be considered in regard to their being integrated them into decision processes of sociopolitical relevance:

- The role of formalization
  As we have seen in the previous section, all of the AI systems require some level of human intervention, which is rarely neutral. Specifically, to address a decision problem one needs to recast it into a form that is appropriate for processing by an algorithm in a computer. In this formalization process, implicit assumptions are made that involve a specific world view. In most cases these assumptions are not critically examined and may contain biases, a priori judgements, or misrepresentations that introduce a slant in the solution that is obtained.

- Data quality.
  When a system is said to be data-driven it is often assumed, without much reflection, that it is closer to reality or somehow more objective. However, data can be incomplete, insufficient, corrupted by noise, or even fabricated. Furthermore, data are also susceptible to manipulation in ways that could be difficult to identify. In practice, the quality of the data is determinant for the quality of the models induced from them. A first condition for the induction of good-quality models is that the data be relevant to the solution of the problem considered, and that no important information is left out. The data should also be as clean as possible and abundant, or, at least, sufficient.

  The preparation of the data for the learning process requires a careful selection and crafting of attributes to characterize the training examples. In some cases, the attributes need to be encoded as numerical values. Since such a quantification requires making further simplifying assumptions, crucial information may be lost in the process. The values of the class labels in the collected reflect past decisions that are not necessarily optimal, satisfactory, or even ethically sound. For instance, a AI-algorithm used to hire workers at a company may tend to select a certain profile, without aiming at diversity, which is one of the desirable characteristics of effective production teams. Furthermore, if the algorithm is trained on data from previous hiring decisions, the system will merely replicate past practices. For instance, it will not be able to identify strategic opportunities in the evolution the company that need to be made possible by changes in the hiring policy. It will also reproduce and possibly amplify biases present in the data; e.g. it will tend to select workers according to their race or gender because of biases in past hiring decisions. Although much research is being devoted to the minimization of such biases, the issue is far from being solved in a fully satisfactory manner.

- Mathematical difficulties.
  The problem of induction of predictive models from data is mathematically ill-defined. Besides minimizing a certain loss function, one needs to impose additional constraints that restrict the space of models that are effectively considered. These inductive biases, which are necessary for learning generalizations (Mitchell, 1980), may take the form of an explicit preference towards simpler models (e.g. the pruning

process is expected to improve the generalization capacity of the decision tree) or the inclusion of regularization terms in the cost function that favor smooth models by penalizing complexity. In any case, these are heuristic strategies that could result in suboptimal models.

- Technological issues
  Even if the type of model were known and the appropriate inductive bias for the problem had been identified, the computational power needed for learning may be excessive.

- Interpretability
  There is often a tradeoff between the quality and the interpretability of models induced from data. For instance, decision trees, which are generally less accurate than other sub-symbolic methods, such as Support Vector Machines, or neural networks, provide a model that can be readily used for human reasoning. By contrast, the high-dimensional embedding carried out in SVMs or the distributed representation achieved in neural networks make it difficult, if not impossible, to gain insight and understanding. In fact, this lack of interpretability is a severe limitation in the deployment of machine-learning models in some important areas, such as medical diagnosis.

- Adaptation
  Another drawback is that data provides a static view of the past and cannot be used to guide evolution. A possible way to is overcome this limitation is to generate synthetic data in model-based simulations. However, in that case, the quality of model becomes crucial. No understanding will be gained beyond what is implicit in the simulated model.

These summarily reviewed shortcomings of a data-driven approach to AI should be sufficient to dispel the illusion of data as an anchor that would guarantee an efficiently managed, fair, objective decision-taking process in any realm of importance to society. The judicious use of data should improve the quality of our models, and, in consequence, of out decisions. However, we need to be aware of the slanted view that data alone can provide and cautiously avoid suspending our judgement.

5. Conclusions

Democracy needs to deal, by its nature, with conflicting choices. Even though AI systems can be used to assist us in the resolution of such conflicts, they cannot be used to avoid our responsibility, as members of a political community, in making such choices. While it is a fact that AI systems are developed on the basis of sound mathematical theories, it in does not follow that for this reason they are neutral, objective, or fair. The use of data in the design of such systems does not necessarily make them more objective, or less susceptible to manipulation and error. On the contrary, there is ample evidence that data-driven models can reproduce and even amplify biases. The quest for fairness in machine learning is an open area of research, one that highlights a severe limitation of current AI systems. More substantial difficulties are encountered in the design systems that can deal with general reasoning problems (Strickland, 2019). Some researchers have called the attention of the AI community to the fact that a more comprehensive approach is needed to address even commonsense reasoning (Davis & Marcus 2015). AI systems, at least at the current stage of development, can, and probably should be used as support tools in the decision process in contemporary democracies. However, they cannot be used to curtail or supplant the citizens' involvement in the democratic process. In particular, they should be explicitly avoided in providing shortcuts to voting processes. Given Searle's (1980) observation that a mere instantiation of a program in a computer cannot exhibit intentionality, it is hard to envision a future in which an AI system can be used to fulfill the central role humans play in the democratic decision process, unless, in Searle's own words, these AI systems do 'duplicate the causal powers of the human brain'.

In summary, conflict cannot be avoided or resolved through technology. As a political body we need to assume the complexity of integrating the disparate goals of individual into collective decisions. No simple solution can be applied, because the existence of this tension constitutes the core of democracy.

6. References

Arrow, Kennet (1963). *Social Choice and Individual Values*. Yale: Yale University Press.

Asimov, Isaac (1955). *Franchise*. Published in 'If: Worlds of Science Fiction' (August issue).

Davis, Ernest and Gary Marcus. 2015. Commonsense reasoning and commonsense knowledge in artificial intelligence. Commun. ACM 58, 9 (August 2015), 92-103.
DOI: 10.1145/2701413

Harari, Yuval Noah (2016). *Homo Deus. A Brief History of Tomorrow*. Barcelona: Penguin.

Mccarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.* Reprinted in AI Magazine 27 (4) Winter 2006
DOI: 10.1609/aimag.v27i4.1904

Mitchell, Tom M. (1980). *The Need for Biases in Learning Generalizations*. Technical Report. Reprinted in "Readings in Machine Learning" (1990) Shavlik, Jude W. and Dietterich, Thomas G. Eds., pp. 184-191".

Newell, Allen, John C. Shaw, and Herbert A. Simon (1959). *Report on a general problem-solving program.* Proceedings of the International Conference on Information Processing, pp. 256-264.

Searle, John. R. (1980) *Minds, brains, and programs*. Behavioral and Brain Sciences, 3 (3):, pp. 417-457
DOI: 10.1017/S0140525X00005756

Silver, David, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, Demis Hassabis. *A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play*. Science, 362 (6419), pp. 1140-1144.
DOI: 10.1126/science.aar6404

Strickland, Eliza (2019) How IBM Watson Overpromised and Underdelivered on AI Health Care After its triumph on Jeopardy! IEEE Spectrum (April 2)

Turing, Alan M. (1950). *Computing machinery and intelligence*. Mind, 59 (236), pp. 433-60.
DOI: 10.1093/mind/LIX.236.433

Winograd, Terry (1971) *Procedures as a Representation for Data in a Computer Program for Understanding Natural Language*, MIT AI Technical Report 235
http://hdl.handle.net/1721.1/7095