



Artificial Intelligence – Big Achievements and Huge Questions Viewed from Mathematics

Cédric Villani

Introduction

Mathematical algorithms have probably been around for more than 4000 years, as suggested by the famous YBC7289 clay tablet (dated 1600 BC or earlier), displaying an amazing computation of $\sqrt{2}$. They have grown in scope, diversity and sophistication together with mathematical sciences. But the middle of the twentieth century marked an amazing new turn. On the one hand, arguably for the first time in history, the outcome of a major human event depended on the devising of a clever, mathematically sophisticated algorithm (this is the story of the work of Alan Turing's team during Second World War). On the other hand, within just a few years, the basis of modern computer technology and computer algorithms were laid with the discovery of transistors and the works of Turing, Shannon, Von Neumann, Wiener and others.

Progress had been slow, but then it accelerated. Fast-forward half a century, and here we are, with a world full of algorithms, and entire sectors of human activities have been revolutionized by algorithms. For instance, to get an idea of how it now looks in world finance, just read the books “6” and “5” by Alexandre Laumonier, providing an impressionistic but thoroughly documented of mysterious algorithms fighting against each other, fortunes evaporating in a fraction of a second, crazy race for speed of transmission and execution. Whether this vision, fascinating and frightening, will extend to all of society, has been the subject of considerable debate; but one sure thing is that algorithms will capture a more and more important role in our economies, our societies, our lives.

A chapter within this long rise of algorithms is the long rise of artificial intelligence. This field is old, by modern standards, since it started almost at the same time as computer science, with the works of Turing and Shannon. Actually, some of the most important methods and algorithms used nowadays in this field did originate from the fifties or even forties. A vision of the founding fathers was that artificial intelligence would help us understand our own intelligence. After some initial fascinating dreams and successes, crystallized in 1968s HAL computer in Kubrick's *Odyssey of space*, the field mostly stalled. Then it accelerated again recently, partly because of new methods, partly by taking advantage of the amazing new speeds and capacities of computers, partly by the exploitation of the huge databases which have appeared. And suddenly Artificial Intelligence has become an enormous hype, with speculations of superhuman intelligence, economic catastrophes; and any ambitious “global entrepreneur” has to keep artificial intelligence under his or her radar. Questions about artificial intelligence and machine learning are so frequent on Quora, appear in broad audience magazines, newspapers; they have also given rise to new directions of research and a renewed attraction for young scientists.

In this context, it is normal to be enthusiastic but to keep away from mystic overhype. It is also normal to remain cautious, and to try and point out questions which remain in the dark. So let us go for a nonexhaustive overview.

Disclaimer: I am not an expert on AI! But the field has been taking so much room that I could not leave it unexamined. Actually I have taken keen interest in the related field of MCMC already for the past 15 years.

1. Basic principles

1.1. Optimization. An intelligent solution is one which tries to find the best analysis, best answer, best action in a certain context. So artificial intelligence will often be about optimizing. Linear optimization, in which the constraints and functions to optimize are all linear, has a rich theory with a lot of structure; but apart from that peculiar setting, not many methods are known for optimization when the setting is rather general.

By far the most popular general method of optimization is gradient descent: follow the gradient. For instance, to find the highest point in a landscape, just look for the direction in which altitude increases fastest, and continue this way. In nature, optimization is supposed to work in a different way, namely through competition (as in natural selection). Parallel to that, there are algorithmic methods based on competition, be it through mutations, auctions or other mechanisms.

Mutations introduce probability theory in the game, and huge progress was made when probabilistic and deterministic methods were mixed: these were, in particular, the Monte Carlo Markov Chain (MCMC) methods, which go back to the forties but have become all the hype in the nineties. Consider again the problem of finding the highest point in a landscape: with the gradient method, you will in general get trapped in a local optimizer. But MCMC can get you out of the trap, by randomly allowing some motions which will get you down, thus leaving a possibility to get to the next hill and in the end to arrive at the true peak. And when arrived at that true highest peak, one will also, from time to time, get down the hill, so that the information is about the time spent in the various states. (And there are techniques to progressively make the exploration deterministic, so that one may eventually settle in the culminating point, or at least a very high peak).

Whatever the technique used, the field of artificial intelligence strongly depends on optimization.

1.2. Learning. But besides the notion of intelligence there is of course not just the notion of finding an optimal, or at least good, response. It should also adapt to the situation, and do things which it was not explicitly told to do. Or, to use a phrase by Samuel (1959), the program should have “the ability to learn without being explicitly programmed”.

The field of machine learning is about letting an algorithm discover by itself a good way to handle a problem, through reviewing information and adapting to that information. One of the very first such systems was Shannon’s electric mouse, Theseus (1952), which would explore a maze to find the best way out.

To continue with the analogy of finding the highest altitude, think that we don’t want to only find the peak, but also to find the shortest path between the starting point and the peak, taking into account what we explored. Or, more ambitiously and more interestingly, that we wish to discover recipes, learnt from examples, that allow us to find the peak very fast, if we are put in a new environment which shares certain features with the previous environments that we explored.

A field of mathematics in which learning has always been at the core is Bayesian statistics. One wishes to evaluate the probability distribution of a certain set of parameters, and for that one starts with a prior distribution, then updates it with all the knowledge gained from successive information.

1.3. Classification. Imagine that you have a number of observations falling in several categories: maybe just two categories, A and B. You would like to describe the difference between A and B in the shortest possible way. In mathematical terms, it could be about separating the phase space in two regions, through an easily described interface, in such a way that all A observations lie on one side, and all B observations lie on the other side.

The best situation is when you can find a hyperplane to do this separation job; by the way there is a long tradition of separating hyperplanes in the context of convexity theory. But of course, most of the time you will not be able to do so. On the other hand, it might be that a change of parameterization gives rise to such a possibility. This is the principle of the method of linear classifiers (so strongly associated to machine learning that an icon about this principle was chosen as the logo of the machine learning Wiki!).

1.4. The curse of dimensionality. In practice the learning problem stumbles across the major problem that the phase space is huge. Already in the simple Theseus problem, the phase space is not the board on which the mouse crawls, but rather the set of all paths in this board, so there is a combinatorial increase of the complexity. But in any realistic problem things are way worse in terms: for combinatorial or complexity questions, problems have to be set in large dimension. Consider the problem of figure recognition: possible variations in shape of written numbers imply that the unknown lives in a space with dozens of dimensions. In some currently used modern models for semantic analysis, language representation occurs in a 300 dimensional space. In phylogenetic reconstruction, the number of possible trees is beyond imagination (500 taxa can be arranged in more than 101275 trees!)

In the absence of specific information to reduce this high dimensionality, there is absolutely no hope to explore the set of possibilities via deterministic, systematic methods. Some guesses have to be made, and one has to resort to chance in a way or another. A good news is that random exploration will give, in many cases, surprisingly efficient methods. A bad news is that it is not really understood why. Another bad news is that there is in general no way to be completely sure that the method will achieve the desired goal.

1.5. The extraction of meaning. There is a classical distinction between information, knowledge and wisdom. How to get the wisdom from the knowledge, and the knowledge from the information, are longstanding multi-faceted questions. Henri Poincaré said it beautifully: An accumulation of facts is no more a science than a heap of stones is a house. The scientist has to order.

But in our current era a no less pressing problem is to extract information from data. Indeed, the amount and size of data, which are considered, makes it impossible to just examine them with human senses and brains.

Some of the most important beliefs underlying current techniques are:

Belief 1. It all boils down to a “reasonable” number of parameters. That means, even if the data we are observing give information about a million sets of measurements in a space of several hundreds of dimensions, eventually we should be able to classify all this information according to a rather small number of parameters, say 10.

To give an example among a multitude, a case which became infamous recently in relation with automated election campaigning, is the OCEAN model (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) which was used to re-present personality traits of users in social media. In this model, all the complexity of behaviours was summarized in a five-dimensional space.

Belief 2. We may let the algorithm discover (with or without precise instructions) these explanatory parameters.

For a long time statisticians have developed methods of “exploratory factor analysis” to identify the determining parameters in a multivariate set of data. Model selection is about finding the good compromise between a too precise and a too rough description of multiparameter problems. On the other hand, some of the new artificial intelligence methods perform such a task by very indirect ways.

While “understanding” is, in a way, about finding the best way to sort things out, this can serve a diversity of purposes, in particular: compress data, recognize data, react to a situation, or generalize data (either by interpolation or by extrapolation).

1.6. Parsimony and extrapolation. Implicit in the previous discussion is the notion of economic representation of the information. This underlying general principle is also made explicit in parsimony theory, in the form of a minimization problem, which has a taste of the entropy problem in statistical physics.

Parsimony could be summarized as follows: given an incomplete set of data, let us find the data, which complements the set in such a way as to achieve minimal “complexity”. There is a certain art in choosing what complexity here means, and it can be based on phenomenology as well as on fundamental principles.

2. Some applications

Applications of artificial intelligence are now legion, and underlie a large part of the current innovation. Here is a very partial list.

- Pattern recognition: e.g. the algorithm will guess that the image which is presented to it is that of a Panda bear or a “priority road” sign or something else, and it will give a “confidence percentage”.
- Prediction: For instance, it is through the sole use of data given by social users networks that the company QuantCube, specialized in data analysis, successfully predicted the election of Donald Trump at the Presidency of the USA, even though with a tiny margin of confidence.
- Preference guesses: from the behaviour of a list of users, and partial information about the behaviour of another user, guess what the latter prefers to see, hear or feel. The most famous example is the “Netflix problem”: given an incomplete table of preferences for a large list of users (this user loved this movie and hated that one, etc.), and an incomplete set of preferences for another user, guess what he or she will love or hate. Accurate preference guesses are a holy grail in the fields of advertising, selling, but also for opinion campaigns.
- Translation: Google Translate rests on statistical methods of machine learning much more than on grammatical analysis.
- Composition: For instance, using artificial intelligence researchers have composed songs in the style of the Beatles, or automatically generated short movies from a selection of images, etc.
- Diagnosis: Expert systems such as IBM’s Watson guess a likely disease based on symptoms, or evaluate a situation based on measurements; similarly, banks detect transactions with a high risk of being identity thefts.
- Clustering and sorting: For analysis of relations between words, languages, species, etc.
- Automation: self-driving cars, drones, etc.
- Interface man-machine: for instance through commands which adjust to the person’s mentality or even thoughts; evolutive prosthetics which learn the morphology of the body.

Etc. Machine learning has become so routinely used, and with such a diversity of tricks, that competitions are now regularly organized between teams to select the most efficient method. The most famous of these platforms is the website [kaggle.com](https://www.kaggle.com)

Implications of these methods are not only in technology: they also suggest new trends to scientists, even though often resting on somewhat shaky ground.

Let me comment on four examples which I came across recently.

- Inria in France has several projects based on robotics control: the motion of the eyes, or even just the thoughts (through recording of the brain activity), would be able to pilot a drone or robot.
- Microsoft Research in Asia developed an algorithm to generate avatars: it uses a collection of pairs (photograph of a face, artist's rendition) to develop its own recipe imitating the style of the artist. Thus when a new photograph is provided, the algorithm will suggest a translation in the style of the artist.
- A biology research paper published in September 2016 suggested that there were four species of giraffe in Africa, rather than one; it was based on learning from large samples of genetic data. (This is an example in which a scientific field is changed by the use of artificial intelligence).
- Riccardo Sabatini and his team showed how to teach a machine-learning algorithm to reconstruct the face of a human from the DNA sample. This is a typical case of application of machine learning: on the one hand the data is absolutely huge, on the other hand the correspondence between the data (genotype), and the expected outcome (phenotype) has famously remained a nightmare for geneticists.

3. The algorithms

3.1. Trends. Machine learning encompasses a diversity of techniques and tricks. Searching online it will be easy to find some lists and selections of them, such as "Top 10 machine learning algorithms" etc. Buzzwords evolve as new algorithms demonstrate their efficiency. A few years ago in lectures on the subject you would hear a lot about SVM (Support Vector Machines); now it is Deep Learning which gets the most credit.

The example of Deep Learning shows that it is important to retain a diversity of methods, and some unconventionality. Indeed, not so long ago, most of the renowned experts in the field would dismiss neural networks as inefficient and doomed; but the tenacity of Yann LeCun demonstrated that these methods can be amazingly efficient.

A beautiful reference about the many facets of artificial intelligence, at least up to a few years ago, is the book by Russell and Norvig. To get some glimpses of the state of the art in current artificial intelligence research, one may watch the online videos of the following emblematic events:

(a) The plenary lecture of Emmanuel Candes at the International Congress of Mathematicians in Seoul (2014), about parsimony methods applied to preference guessing as well as medical imagery; themes such as the right definition of "complexity" of an image, or the mathematical justification of the method, are enlightening.

(b) The course by LeCun at Collège de France in Paris, and the seminars given there, e.g. by Ollivier and Mallat, which will provide a diverse view as well as many questions.

A general remark is that, eventually, artificial intelligence algorithms boil down to technical keywords such as large matrix diagonalization, convex optimization, gradient flows etc. which to an outsider hardly evoke anything related to "intelligence".

3.2. Neural networks. Neural networks is just one of many fields in Artificial Intelligence. But it has become such a craze that it deserves a specific review. Let me just mention that

- Neural networks use a list of examples which we may write (x_i, h_i) , and the goal is to produce a "rather simple" function h in such a way that $h(x_i) \approx h_i$; in other words it is about guess an unknown function through examples;
- Neural networks are made of nodes (neurons) and links (synapses); while the general pattern is inspired from animal brains, the organization is quite different; nowadays the neurons are organized in a rather large number of layers (depth), with synapses joining neurons from one layer to the next one;
- Each synapse corresponds to an elementary nonlinear function, approximating a step function, and depending on some parameters; this mimics the fact that a synapse can transmit more or less information, and does so only at a certain level of excitation; so a function is a combination of elementary nonlinear functions;
- The number of neurons can be very large nowadays, with millions of parameters, and actually large neuron networks have achieved amazing results these past few years;
- The optimization procedure is based on a gradient flow method, here called "back-propagation".

I refer to the lecture of Yann LeCun for more information about networks and their use. I also refer to the lecture of Demis Hassabis for an in-depth discussion of the spectacular and instructive case of the AlphaGo program,

which achieved world fame by demonstrating its super-human level at Go and showing at the same time what could be considered as creativity.

It is important to point out some conceptual differences between AlphaGo and a “classical” algorithmic approach to Go playing. AlphaGo achieved its super-human power by an extraordinarily intense training, taking primary examples from recorded Go games by humans. The particular rules which AlphaGo applies do depend on the particular examples that it was fed. But also, AlphaGo spent an enormous amount of time playing against himself and trying random departures from the sets of games it was given. Thus a good amount of randomness enters the making of AlphaGo, first through the selection of games it is primarily fed with, secondly through these random variations that it is trying.

4. Big Scientific Questions

The brilliance of programs such as AlphaGo, or the ever-increasing number of applications and programs which use modern AI, demonstrate the impact of these methods. But still this comes with big questions.

4.1. Why does it work so well? This question, which is formulated verbatim in Mallat’s contribution at the Collège de France, is on the lips of every researcher, especially since the surprise comeback of deep neuron networks. In fact, these methods are vexingly efficient, and took theoreticians off guard. For sure there is, among other things, an effect of the “Big” factor. It has been noted already some time ago that the most important asset of a database is its size; inaccuracies being washed out by the sheer number. With modern methods we also see that size does matter.

As Ollivier emphasizes in his own contribution, to better understand this question there is need for a much more conceptual modelling, with a geometric study of the phase space and the process.

A related question is: Which problems can be solved? Mallat likes to formulate this in term of three keywords which are well known to harmonic analysts: complexity, regularity, and approximation theorems. Working in the particular context of parsimony methods, Candes insists on three ingredients for success which seem important: (i) structured solutions (the fact that only a few parameters really count; mathematically speaking this would be, typically, about a matrix having small rank, or being very close to having small rank); (ii) the ability to use convex programming for computational purposes; (iii) incoherence, that is, the fact that the missing information does not present any particular pattern in respect to the key parameters. With a proper mathematical formalization of these assumptions, Candes and Tao are able to prove a few neat mathematical theorems showing accuracy of the reconstruction for “most” samples.

Yet another related question, obviously, is: can one make those algorithms more efficient? There is real motivation for this, as modern algorithms are, by any rate, inefficient: they are very demanding in terms of storage, go through datasets dozens or hundreds of times, use up absurdly enormous power if we compare them to a human brain, do not yet adapt to quantum algorithmics...

Part of this inefficiency is also due to the use of randomness, which is typical (randomness is inefficient but in complex phase spaces all other methods are usually worse). Arguably, it may also be due to the poor incorporation of rules and semantics in the search for representation.

4.2. MCMC methods. I would like to recall that MCMC methods were all the buzz in the 1990s to magically solve problems with large phase spaces. The articles by Persi Diaconis on this MCMC Revolution are very instructive.

Arguably, MCMC was a particular case of machine learning, with a modelling (the probability distribution) which was improving with the amount of data, using randomness and gradient flow optimization.

But the analogy does not stop here: some of the same questions as above were also central: Why does it work so well? Which are the geometrical or structural conditions which make it work well, and so on?

Diaconis, who became famous for his discovery of the cut-off effect in the convergence of Markov chains (that is, the fact that the convergence often occurs very rapidly after a certain time, going in a small number of iterations to “hardly mixed” to “very much mixed”), has been fascinated by the problem of mathematically explaining the efficiency of MCMC methods. Together with Michel Lebeau and other collaborators, he worked on analysing this in controlled cases with very simple rules. The results, which appeared on the prestigious journal *Inventiones*, are fascinating: even if the model is oversimplified, they are based on an amazing level of mathematical sophistication, and the convergence estimates are quite conservative. While these authors have established admirable pioneer work, it is likely that there is still room for huge improvements.

By analogy, the following question is very natural: Is there a sharp cut-off effect for AI algorithms, in terms of the size of the data, or the number of parameters?

4.3. What about our intelligence? The dream of the founding fathers was that artificial intelligence would lead us to a better understanding of our own intelligence. So far this has not borne so much fruit. On the contrary, some striking experiments suggest that our algorithms are very different from those which are used in AI. Maybe none is more spectacular than the correlated noise attacks performed by Christian Szegedy: from an image which is clearly recognized by the algorithm (say a truck), a tiny modification (invisible to a human mind) will fool the algorithm into recognizing an ostrich, with extremely high confidence.

Even without this, the inefficiency of artificial algorithms with respect to natural ones has been an elephant in the room: just a few observations are sufficient for a human to identify a pattern, where machine learning algorithms need huge numbers of them.

In such situations, however, comparison between natural and artificial mechanisms has helped suggesting new research directions, such as reinforcement by adversarial training (see LeCun's lecture), or the modelling of universes with categories and subcategories (see Tenenbaum's lecture).

Also, at qualitative level, some striking suggestions have been made by Dehaene, for instance about the encoding of numbers in the brain, based on artificial neuron networks.

It is likely that these features (strong discrepancy between natural and artificial algorithms, but mutual influence in their understanding) will continue, and that little by little we shall identify some ways to model some human intelligence features through AI.

4.4. Epistemological questions. Will AI be able one day to do science, to out-perform human scientists, or, more modestly, to help humans finding new science models, or science laws?

Mathematics has been a favourite science in this question, probably because (a) mathematics does not explicitly rely on experiments, (b) mathematics is the only science in which the rules of the game are fully known, (c) mathematics is both familiar to and admired by the (mostly geek-type) conceivers of AI programs. So the idea of a theorem-proving AI is a widely shared dream in AI.

Besides mathematics, one could hope for AI to identify patterns, formulas, or even equations, without "proving" them, but showing that this is how nature works.

One may object that mathematical proof requires exploration of an extraordinary combinatorics. Automated proof checking has gone a long way forward, but there is a whole world between proof checking and proof making.

One may also object that machine-learning methods, based on examples rather than models, will be poor at discovering new laws and reasons. But on the other hand, we have examples in science of laws which were first discovered through the examination of data and later turned into laws: one of the most famous is Kepler's law of elliptical orbits.

Still, so far the harvest is meagre. It is true that a computer program has been good at deriving the basic laws of Hamiltonian mechanics, and that some expert systems have managed to prove some nontrivial geometry theorems, but the whole of such achievements remains a tiny portion of science, and I am not aware of any novel law which has been found through AI. Let me also bring back the spectre of MCMC by recalling that it was originally used to discover the phenomenon of hard spheres transition; but that nobody has been able to justify or understand this phenomenon in more than half a century.

For the moment we may just say that time will tell!?

Now, one may for sure be more optimistic in the prospect of an AI-aided scientist, and there are already such examples, especially in biology. In his seminar, Mallat also shows how to use AI to derive the shape of an unknown energy in a mathematical physics problem.

However, these are not really about finding new laws, but about finding new ways to organize a complex given information. Here, for sure the most important themes revolve around genetics and related fields, such as phylogenetics and taxonomy.

An example of progress in the field of taxonomy is the recent work on giraffe genomes, which suggests that there are actually four species of giraffes (note that the notion of giraffe is no longer clearly defined!). As for the field of phylogenetics, which aims at identifying the "parenthood" relations between species, a recently debated issue was the respective places of Archaea, Bacteria and Eukaryotes. In these fields MCMC and other machine learning methods have been used on large genomic samples. This is exciting, but leaves some big questions.

A first big question is how will researchers master these tools, and most importantly the safety rules for their use. Thinking again about MCMC, there is a well-known course by Alan Sokal warning users that results obtained

by MCMC have no scientific value whatsoever if they do not come with justification of the convergence times and sampling rates through an estimate of features such as the autocorrelation times. It is likely that most of the published scientific literature based on these techniques does not perform such checks. Of course debates follow.

A second one is about the epistemological status of advances which have been obtained through AI algorithms. There is usually no proof of convergence of such algorithms, and thus no way to guarantee the accuracy of the method. Should we admit them as evidence, knowing that they use randomness and other black box features?

A third big question is about the meaning of “understanding”. AI methods have made huge progress when we became more lenient in our demands for understanding the rules which produce the results. The good thing is that the algorithm does usually much better than what we could imagine, but the bad thing is that we don’t understand the reasons for the outcome, even when we have it. To remedy this, one should work (and one already works) on the way to display and propose the results, singling out those parameters which played a most significant role in arriving at the result.

A final big question is the risk of seeing drops in the mastering of entire chunks of scientific skills, namely in the modelling. For instance, in mathematical finance, stochastic modelling is rapidly giving way to big data analysis. One certainly should rejoice about this diversification, but one can also worry that stochastic finance analysis, based on modelling, may soon be forgotten by younger generations of finance mathematicians. The same can be said about many fields. Whatever point of view one wishes to adopt, it is important to recall that in a classical scientific view, understanding always includes modelling, and it is certainly foolish to believe that data will get rid of that. Just think of the big difference between cause and correlation, that only a model can bring!

5. Big Societal questions

AI methods have invaded most fields of technology and will very likely be used more and more, for more and more tasks. This is a partly comforting, partly worrying trend.

5.1. How robust is AI? Szegedy’s experiments have shown that AI-based recognition may be fragile, and possibly subject to attacks exploiting the fact that it has been trained in a certain way. Currently, AI remains so far good for specialized tasks (like playing Go!) and this may lead to a lack of stability and robustness.

It is notable that one of the most promising directions of research in AI, namely adversarial learning, is precisely aimed at making learning algorithms improve by challenging them in situations of ambiguity (as when one is training a youngster by giving exercises with traps).

In the case of AlphaGo, Lee Sedok was able to fool the algorithm once by leading it into a highly non-comfortable zone that it had not explored enough. This is reminiscent of human strategies against chess programs. It certainly would be impossible with the current version of AlphaGo, which is way stronger than the one which Sedok played against. But it shows that it is not easy to ascertain the robustness of AI.

5.2. Who will take responsibility? The achievements of AI are impressive, but the convergence is not guaranteed, the mechanisms remain mysterious. Who will take responsibility in case of legal battle or policy change? The question may be asked for an automatic driving car, but also in a number of other situations. For instance, it is now possible, through AI, to get a rather good reconstruction of the face through DNA sample: can this be used in a criminal case?

Then, there are discriminations which may come in AI programs. What happens if an AI program automatically leads to different rules and behaviours in front of different ethnical groups, or social groups? Such situations have already occurred.

5.3. Biases. The case of Tay, the chatbot by Microsoft, has been on the media: that AI was influenced and manipulated by users which transformed it quickly into a horrible racist (and, by the way, a worshipper of Donald Trump). This shows that AI, like humans, may inherit strong biases from their environment, impairing their tasks.

Currently, the method of building an AI, through example-intensive machine learning, leads to biases: the program will depend on the databases which we use. A 2016 paper by Caliskan-Islam, Bryson and Narayanan was pointing that Semantics derived automatically from language corpora necessarily contain human biases.

5.4. Myths and fears. AI has triggered a number of myths and fears already. One is the emergence of a super-human intelligence. The way some authors talk about it makes it closer to religion than science. By the way, this was the subject of a terrible movie, *Singularity* (translated in French by *Transcendence*, which had a very clear christic analogy). It is certainly a good advice to try and keep being objective.

Another fear is that humans will lose their thinking abilities relying too much on AI. Actually there has always been a transfer of tasks from humans to technology as the latter improves (for instance we are not so keen now

about memorizing or computing since our books and computers do it so well for us). This debate was already going on at the time of Socrates with the technology of writing.

On the other hand, new technologies also come with new challenges, new ways to entertain. Computers have taken some abilities from humans, but also brought new abilities!

Also, a striking case in point about the AlphaGo experiment is that the new supremacy of algorithms on humans did not seem to deter humans from playing Go; actually, as Hassabis pointed out, the sales of Go games skyrocketed as a consequence of the competition. Playing with humans is a human activity after all.

5.5. Economics. In the field of economics and collective social affairs, things may be more worrying. So far the two main areas of concern are (a) robotization which may, to some extent and for some time, deprive humans of jobs, and (b) over personalization of interactions which may create bubbles and weaken the cohesion of society. Let us briefly discuss both.

The replacement of humans by robots and algorithms in certain tasks is a robust trend. In the case of AI it typically concerns medium skills (not the lowest, not the highest). As one example among many, in 2016 Foxconn has announced the replacement of 60,000 factory workers with robots.

The replacement of jobs may take place at the level of production sites (factories, companies) but it may also go in the interaction between humans and the task. Famous cases are those of photograph developers or travel agencies: those have mostly gone now, and users are handling it themselves with the technology. The same may occur with AI. Bankers, taxi drivers, generalist doctors, are all categories for which speculations are going on. Short-time traders are a famous category which was upset by algorithms, and there is hardly any doubt that finance will rely more and more on AI, jeopardizing classically trained finance officers.

In a Schumpeterian view, one may argue that these jobs will reappear in another format. But the characteristics of the present wave (so strong, so versatile, so quick, so globalized) make it possibly different from the previous ones; it may be that the “creative destruction” takes quite a long time by human life standards. The positive part of it is that there will be a whole chunk of new, high-level jobs devoted to interpreting, mastering or accompanying AI; so a bright future opens up for related jobs. In this context, statistician was elected CareerCast’s “Best Job of 2016”.

The second main concern is about the over efficiency of algorithms in personalizing the interaction. A few years ago the public diffuse fear with algorithms was mostly about a systematic uniform treatment for everybody, and it is ironic that now it is the other extreme which arises fears.

While personalization is certainly good news in certain fields (personal medicine for instance), it may be a problem in respect to solidarity. Insurance is about sharing risks, but if risks become tailor-made for each individual, the solidarity may effectively disappear. Profiled news will get citizens exactly the information which they wish to hear. Profiled politics will get the politicians to adapt their speech and convictions to exactly what their electors wish to receive; and it will help their campaign teams manipulating their opinions. Profiled suggestions will at the same time keep customers in their preferences and satisfy them (it was recently a sensation when it was found that Netflix used some 80,000 subcategories of users). Profiled advertisement will help trick customers into buying (while already trying to cover up the profiling by inserting some more random advertisement). All this is so efficient that it may have a destabilizing effect in a number of human affairs. Actually there is already ample clue that Big Data and AI methods played a significant role in both the controversial Brexit campaign and the controversial Trump campaign. In the same way as the narrative of Internet has been switching from freedom to mass surveillance and from sharing to bubble creation, it is possible that the narrative of AI will switch from fine support to manipulation. Already a notable book has appeared by the mathematician O’Neil with a strong title that says it all: *Weapons of Math Destruction: How Big Data increases inequality and threatens democracy*. This evolution, in conjunction with uncontrolled phenomena of fake news, “trolling”, distortions and the like, may possibly be one of the most important problems facing humanity currently.

Let us also note that these topics are still controversial (as shown by Mark Zuckerberg’s recent statement in disbelief of bubbles), that the environment is rapidly evolving, and that experiments are virtually impossible to do; so that it is not clear if the subject is amenable to science.

5.6. Human-AI interaction. One of the most fascinating features with AI is the interaction between humans and algorithms. This comes together with issues of human-human interactions.

It was already noted that human-AI matches do not deter human-human games. In medicine, algorithms may soon be able to outdo humans in diagnosis, but the patient-doctor relation is also one of human trust, and patients are certainly not ready to confide their emotions and fears to an algorithm (even though, on the contrary, some humans are much more comfortable to confide their traumatic experiments to neutral robots than to

fellow humans). Because of this, or thanks to this, the paradigm of the medical doctor using AI is bound to be much more powerful than just the medical doctor or just AI.

Related to trust are subjects of responsibility and explanation procedure. It has already been widely debated that the responsibility issue for automatic car driving may be quite tricky. Also, even though automatic driving will certainly be more secure than human driving, some people object to, or fear putting their lives in the “hands” of an algorithm. In this case the job of driver cannot be considered similarly as the job of medical doctor: first because the affective bond between passenger and driver is much weaker than the bond between patient and doctor, but also because a human driver is very poor at securizing an automatic car (for taking the sequel of an automated procedure, or intervening in an emergency procedure, we humans are very bad). So the automatic car will basically have to be fully automated.

As a different example, consider the problem of detection of frauds by identity thefts. Already in the nineties, major companies were using insurance procedures and a certain number of rules to refuse transactions which were considered “fishy”. In the absence of any explanation and any interaction, these gave rise to infuriating incidents. Nowadays banks are equipping themselves with AI-based recognition algorithms for what is fishy and what is not. When this is well designed, this comes with human arbitrage, explanation, and reaching out to the customer (through cell phone, for instance), so that responsibility is clear and interaction with the customer can take place. Of course budgetary issues, efficiency, trust to the consumer, and so on, will also be elements of choice for a bank company willing to improve in this direction.

It is certainly an interesting multidisciplinary subject to understand when the human-AI combination is an improvement and when it has to be fully human, or fully AI.

In any case, in such a context, the education of a wide audience to the basic principles of AI, with their powers and limitations, seems like a wise society option.

6. Bibliography

The bible of AI is the beautiful book by Stuart Russell and Peter Norvig, *Artificial Intelligence, A modern approach*. Even if it does not discuss so much the most modern algorithms, it is an amazing synthesis and a work of art.

Here are some interesting books about AI and its interaction with society: *Probably Approximately Correct* by Leslie Valiant (a milestone in the algorithmic approach to AI); *The Technological Singularity* by Murray Shanahan; *The Future of Machine Intelligence*, by David Beyer.