Teaching computer science to legal professionals

Serge Abiteboul,
ENS, INRIA, Paris;
&
Claire Mathieu,
CNRS, Collège de France

Abstract:
The education of legal professionals in the 21st century should have a significant scientific dimension encompassing the information sciences and digital technology.

An attorney's nightmare

Legal professionals in the middle of the 21st century are constantly using information and communications technology (ICT) and digital tools. Their library is organized in a distributed database, the librarian is an automat that not only provides the most relevant information without the professionals even having to say what they want but also writes reports in their stead. The client might be a software program in bankruptcy owing to the accused party, a malicious piece of software. The judge of digital tort is a software program certified by an algorithm. We might suspect that lawyers will feel they are second-stringers in this digital universe, but that is not at all so because they, too, are machines. When the legal professional realizes this, he wakes up. A nightmare about the realm of justice, where dehumanization has denatured the profession, a world where anything human is redundant and should be adapted to an implacably indifferent environment.[1]

Why do lawyers have to learn information science?

The information and computer sciences have set off a revolution in justice systems. Rather than leaving, in vain, justice petrified (operating as if in the 20th century) or letting others make choices about the orientation for our society, we must collectively choose how to use digital technology to improve justice. In this context, citizens, above all jurists, have to learn how to handle these new tools. They have to have a command over the concepts in ICT in order to choose the changes to come rather than passively accept them.

What must legal professionals must in the digital sciences? The answer forces us to imagine how they will be working in the future. Let us rapidly look at the professionals who will be working in legaltech startups or developing software for corporate attorneys in firs of various sorts. Since their work lies at the junction of two fields, they will have dual qualifications, in the law and in ICT, that they acquired as part of their university education. As for the others, the mass, they will have to acquire the basics that will enable them, like other citizens, to live in a digital world. In addition, they will have to have more knowledge in order to understand the tools to be used in their profession or the affairs to be judged.

[1] This article has been translated from French by Noal Mellott (Omaha Beach, France).

Let us take the example of an accident with a driverless vehicle. Who is liable? There are several possibilities: the company that designed the vehicle, the company that made the software that drives it, the company that supplied the brakes that did not respond in time, the persons who made the library of the piece of software that was not updated in time, those who conducted machine learning, the user who gave the order to drive at night on an icy road…

Another example: the online platform (APB in France) for enrolling applicants to courses of study in institutions of higher education. It is mathematically impossible to satisfy all conditions if the law requires accepting applicants for a course of study for which the number of places is less than the number of applicants and if the law allows neither for examining secondary school records (in order to select among applicants) nor for drawing applicants at random — thus with no way to reduce the excess number of applicants. To keep the law from making it impossible to find legal settlements for problems, jurists will have to have an understanding of the limits of automated solutions: what is impossible as well as what is possible but unfeasible because it cannot be done in a reasonable time. Otherwise, decisions based on the law will be inapplicable. However, if there are not enough safeguards, private firms will rush to develop software to take advantage of the situation and optimize their profits without any consideration for the commonweal.

To balance these considerations, jurists will have to understand the technical or scientific questions in which they are not normally competent (even less so when they have received a traditional education in the humanities). Alongside this pillar of classical education, the digital revolution has set up a new pillar: the need for solid scientific knowledge.

What should legal professionals study?

In this new world, there will be an abundance of digital issues with which jurists will have to deal. Their education will have to prepare them for this by providing training in the information and computer sciences. What skills will they have to learn?

To form a legal opinion on questions involving algorithms, it is necessary to have basic knowledge in algorithms. For a legal opinion about questions related to accessing and using the Internet, it is necessary to have knowledge about networks and machine learning. To understand what makes a computer run, it is necessary to know about programming, computer architecture and databases. Legal experts in a dominantly digital world will have to have basic knowledge — a sort of abridged bachelor's degree — in the information and computer sciences.

Within the next few years, secondary education will hopefully evolve so that law school applicants will have already acquired basic knowledge in these sciences. They will then take university courses more in line with justice and the law. To take an example: knowledge in security issues is needed to determine liability in a case of usurpation of identity. To understand digital authentication, legal professionals will have to know about cryptography. Among the other subjects to be learned are search engines, data analysis, big data, blockchains, smart contracts…. These professionals will have the opportunity to apply their knowledge by realizing projects in which they have to code "realistic" software.

The law and the code

Computer programing involves, very much like the law, applying rules. In the computer and information sciences, there is also a hierarchy of the programming language's degree of precision, similar to what exists in the realm of justice. An algorithm is a law; the software program is the enabling order; and the translation into machine language on a given computer is the circular. The lower bounds and the results in relation to the complexity of problems form a sort of natural law.

Computer programming and the law do not exist in separate, parallel worlds. They manage to find each other. Interactions between the code and the law have been analyzed.[2] Code can "implement" the law, as happens in smart contracts. It does not replace the law but must, instead, comply with it. Society and state have the role of seeing to this, of verifying whether the law has been correctly interpreted.

Software programs are ever more frequently used in justice, where they already perform repetitive tasks and help cut costs thanks to dematerialization and automation. Lawyers are using them ever more often to examine case law. The number of sales contracts managed, even arbitrated, by software is increasing. New algorithms are being made to assist judges (by, for example, assessing the risk of a convict committing a second offence) or litigants and their lawyers (by, for instance, predicting a lawsuit's outcome).

In the United States, an algorithm was installed to help judges make better informed decisions about the conditional discharge of potential second offenders, but the recidivism risk calculated for individuals proved biased. A comparison of the *a priori* assessment of this risk by race (African-American or white) with the *a posteriori* statistics (whether or not the person committed a second offence) showed that second offenders received an *a priori* risk assessment that was, on the average, higher when they were black; and the same held for persons who were not second offenders. By mathematically formalizing the concept of "risk score" and the conditions that this score should meet in order to be deemed fair, Kleinberg, Mullainathan and Raghavan have demonstrated that fairness is impossible.[3]

When the aforementioned APB program had to choose (while lacking information that could have helped make a selection) secondary school students who requested the same course of study where the number of applicants was superior to the number of places available, the problem as posed was impossible to solve. The only solution: choose winners randomly, a method of selection that would determine the winner's career in higher education and afterwards. Such an approach fell in line with the French tradition since, in the 19th century, sortition was used to determine the duration of conscription (from one to three years) of young men. In the 21st century however, this procedure is felt to be unfair. Although Themis, the ancient goddess of justice, was blindfolded, human beings no longer tolerate the idea that dark forces outside their control can decide their fate. Justice in the sense of 21st-century France requires that the unlucky applicants can obtain explanations about why they have been refused admission. Losers demand that the decision be justified. When algorithms make important decisions in our lives, we expect them to be accountable, to be able to explain their decisions. This is especially true in the realm of justice.

Providing an explanation has another advantage, since it is indispensable to verify that the results are not biased. A software program that consults a database and informs a legal professional of cases similar to the one he is handling must not skip over past cases that could be relevant. Skipping such cases carries the risk that the professional's view will be skewed, perhaps by making him less optimistic than he would have been, by hiding possibilities for setting the dispute, or by presenting him with a set of past court decisions that, too uniform, reins in his imagination.

---

[2] LESSIG L. (2000) "Code is law: On liberty in cyberspace", *Harvard Magazine*, January/February, 5p.

[3] KLEINBERG J., MULLAINATHAN S. & RAGHAVAN M. (2016) "Inherent trade-offs in the fair determination of risk scores" available via https://arxiv.org/abs/1609.05807.

Governance by data

An interesting facet of the digital transition in general and especially in the realm of justice is the use of big data to advise, evaluate and optimize human beings. It is rather easy to collect masses of data, and to aggregate, share and analyze them. The digital services installed in law offices often have the goal of retrieving data in order to assess the service's quality and improve it. This seems evident.

But for services of justice performed by human beings, we have many more qualms about such assessments. Nonetheless, the latter can be made with the help of digital data. This could provide an original method for improving justice by, for example, evaluating the biases of judges. The "governance by data" might displease persons whose quality of work will come under scrutiny, but it could have a part to play like the assessments of the costs of a doctor's decisions — not as a binding restriction but as a means for drawing a judge's attention to possible biases. If obsessed with questions of consistency, a government might, of course, be tempted to use these analytical tools to pressure judges to be more efficient or to cut costs. Knowledge of statistics and data science, and of their limits, could help us resist this deviation.

Fantasies and reality

Software programs assist human beings. Unlike mathematical formulas however, they help with processes that are too long to be done by hand. Since we cannot follow them step by step, they have a dimension that eludes us. They are enshrouded in a cloud of mystery, and we tend to attribute to them powers they do not have.

Nowadays, algorithms can help judges and attorneys only in very special contexts, with very simple, finely determined tasks that are not of vital importance (*e.g.*, managing petty offenses or commercial disputes) but always under human control. They are far from capable of replacing people or, even more, of handling the full complexity of serious cases in the realm of justice.

The data manipulated by algorithms are not physical measurements. They are human data with a load of errors, imprecision, incompleteness and contradictions that machines have a hard time handling. The processing of this sort of data is often based on machine learning, in other words on a mimicry of human choices. Machine learning calls for a huge quantity of examples (big data). It works poorly when data are scarce, or the situation being analyzed is unique or atypical. An education, even elementary, would demystify the results. The progress being made in digital technology is so rapid and impressive (even for computer scientists) that algorithms might someday manage to deal with such situations, but this is not so today.

From the perspective of technological progress however, we should ask a few questions. Among the tasks that we now do ourselves, which ones do we want to delegate to machines either without any human control or in a limited way under human control? Which ones must remain human? How to learn to work with algorithms?