Humans and artificial intelligence face-to-face:
The ethics of relations
between humans and machines

Paul-Olivier Gibert, CEO, Digital and Ethics

Abstract:
Human beings are being placed in the position of delegating actions to autonomous machines. This rupture in the relationship between people and machines raises ethical problems. What should (and can) be the "right" use of the new possibilities offered by machines? How to fulfil the dreams that some persons have of artificial intelligence while avoiding the nightmares that others dread? Autonomous machines, reputedly intelligent, will have to incorporate — in their very design — ethical rules and restraints so that the aforementioned delegation be made under the conditions to be expected. In parallel, people's capacity to judge and act on their own and with full awareness must be developed so that they not be subjugated by the machines they have built. We cannot abdicate our free will.

In this second decade of the 21st century, artificial intelligence (AI) seems to embody the Promethean anxiety stemming from technical and scientific progress.[1] Once developed by human beings, it will, according to Stephen Hawking, "*take off on its own, and redesign itself at an ever increasing rate […] Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded.*"[2]

Can AI overtake humanity, its inventor? Some dread this while others, in particular those enthused with transhumanism, are hoping that AI, combined with other forms of technology, will allow us to move beyond the human condition. Between these two extremes is a wide spread of other positions. The media have made ample room for discussions about the hopes and fears aroused by AI. Nonetheless, we are forced to admit that practical, concrete applications are still far from the enablement of "intelligent agents" to fully rival human beings. Machines (algorithms run by computers) can beat human champions at chess, go or poker, but these are very limited human activities.

However we now observe a break (breakthrough or breach?) due to AI in the relations between people and machines. The use of natural language, in particular the spoken word, to communicate with self-learning machines places humans in the position of delegating actions to autonomous machines. This break calls for thought on two ethical issues. On the one hand, what are, or should be, the "right" uses of this new possibility offered by machines? On the other hand, how can we realize the hopes placed in AI while deflecting the dreaded nightmares?

---

[1] This article, including quotations from French sources, has been translated from French by Noal Mellott (Omaha Beach, France). The translation into English has, with the editor's approval, completed references.
[2] Cellan-Jones, R. (2 December 2014) "Stephen Hawking warns artificial intelligence could end mankind", BBC News, posted at: http://www.bbc.com/news/technology-30290540.

The delegation of decision-making to machines: Ethical questions about our relations with machines

AI is a vast, heterogeneous set of quite different techniques for simulating intelligence. It refers, in some cases, to computers with an architecture that reproduces the structure of the brain (neural networks)[3] for the purpose of improving performance. It sometimes refers to simulations via interfaces, in particular communications in the natural language of everyday life or even the technology for voice recognition and reproducion, such as Siri and Google Now. AI's simulation is sometimes itself simulated: "*Designing code that seems smart has nothing to do with a machine that learns all alone. You set rules for it and program the response that the machine has to have if it encounters different types of cases. This is what happens in 90% of what is being sold as artificial intelligence.*"[4]

Regardless of the techniques used, the break between people and machines is quite real: we can delegate to a computer actions by talking to it like a human being — without knowing beforehand how the computer will perform the actions. Owing to the progress accomplished in the automatic processing of natural language, computers can interpret the language that we use in order to translate it into instructions that they can process. This trend, now visible in the form of the virtual assistants on smartphones (Siri or Google Now, for instance), is highly symbolic. The Turing test, imagined by Alan Turing in 1950 to tell human reasoning apart from machine computations, is based on the ability of computers to reproduce the semantics of a human conversation.[5] Year after year, the machine's ability to dialog has grown, and has buttressed the idea of a rivalry between computers and people.

This break in man-machine relations has raised questions about the right ways to use this new technology: "*This week, hundreds of engineers and academics signed an open letter to warn about the abuses of a form of artificial intelligence (AI) that is making technical progress without any consideration of the societal progress that it should bring.*"[6]

Now equipped with more powerful computers and methods for processing data (especially big data), AI is developing tools with a capacity for learning and operating autonomously. This breakthrough signals a break in our relations with machines.

Let us simplify. Till now, there have been four phases in the relationship between people and machines, each phase corresponding to a type of machine:

● Hand tools produce an action only if human beings use them directly.

● Machines using external resources (energy or computational power) perform the tasks that they are expressly directed to do; they can continue operating owing to inertia.

● Machines using external resources (energy or computational power), perform the tasks for which they have been programmed; they can continue operating owing to inertia.

● Machines acquire information and learn, a capacity that enables them to accomplish tasks for which they have not been expressly programmed.

The last phase corresponds to what AI makes possible. It signals a deep break, a rupture, in man-machine relations. During this phase, owing to these autonomous machines, people will confer decision-making tasks on machines (computers of the third generation). Thanks to AI, several digital devices now allow for delegating actions that entail some form of decision-making.

[3] BLAYO F. & VERLEYSEN M. (1996) *Les Réseaux de neurones artificiels* (Paris: Presses Universitaires de France, Collection Que sais-je n° 30421).

[4] GOURITIN T. (5 November 2017) "L'arnaque chatbots durera-t-elle encore longtemps" posted on: https://www.frenchweb.fr/larnaque-chatbots-durera-t-elle-encore-longtemps/305697.

[5] TURING A. (1950) "Computing machinery and intelligence", *Mind*, 49, pp.433-460. Available at: https://www.csee.umbc.edu/courses/471/papers/turing.pdf.

[6] CHAMPEAU G. (14 January 2015) "Lettre ouverte sur l'intelligence artificielle (IA) et sa place dans la société" posted on: http://www.numerama.com/magazine/31868%2Dlettre%2Douverte%2Dsur%2Dl%2Dintelligence%2Dartificielle%2Dia%2Det%2Dsa%2Dplace%2Ddans%2Dla%2Dsociete.html.

In the relationship with machines, we are drawing closer to the situation that exists among human beings, namely: granting a proxy the power to act in one's behalf. In the eyes of the law, this delegation of a power corresponds to a situation where one person (for example, a CEO) confers the right to exercise a power on someone else (a staff-member), who will exercise it under the CEO's responsibility. In other words, to delegate is to confer on someone else the power to make a decision in one's stead. This is far from the simple case of giving an instruction to someone who executes it.

Navigation systems currently provide a striking example of this. The GPS, in particular, takes into account the current flow of traffic to propose trajectories that minimize the time or distance for a trip. The GPS tracker has overtaken human abilities (somewhat as in the game of go) since it integrates in real time information (about traffic jams, etc.) that lies beyond the human driver's field of perception.

The next step in this process of delegation is self-driving vehicles that make all decisions about a vehicle's movements, even in critical situations such as accidents. Among the proposals for coping with the consequential ethical dilemma, we might mention the "*moral machine, a platform for gathering a human perspective on moral decisions made by machine intelligence*".[7] It compiles human viewpoints about the moral decisions that autonomous machines should adopt in such situations.

The quest for the "right" delegation of actions to machines

Ethics is relevant for identifying the rules that should govern the behavior of autonomous machines. Since this technology is recent, and its realizations are still mostly experimental, a pragmatic approach is warranted.

Figure 1: Acting "right": Ethics (thinking, questioning), morals (fixed rules, value systems), law and deontology (laws, regulations)

What does an ethical approach bring compared with legal, moral or deontological approaches? These fields are complementary (*cf*. Figure 1). Law and deontology rely on normative rules with which everyone (in the case of law) or all the members of a profession (codes of fair conduct, professional standards, etc.) have to comply. Morals refer to established rules stemming from value systems (possibly religious, such as Christian morals). Ethics relies on reasoning, on questioning the situation under examination. An ethical approach is useful when the other

---

[7] http://moralmachine.mit.edu/

approaches have trouble answering the question about the right use of an object or the right behavior in a given situation. An ethical approach is well suited for responding to questions about the right use of autonomous AI machines, given the diversity of use cases. When using this technology in my professional activities, I ask four types of questions related to ethics:

- What assignment is being conferred on the machine?
- How reliable are the results produced by the machine?
- What is the information of individuals?
- What consent have individuals given?

To illustrate the application of these questions, I have excerpted from an article passages about fields where AI is expected to have a major impact.[8] These use cases shed light on the socioeconomic issues and ethical questions.

"*AI is completely reshaping life sciences, medicine, and healthcare as an industry. Innovations in AI are advancing the future of precision medicine and population health management in unbelievable ways.*" What are AI's objectives: to provide care to individuals or to address public health issues? If the latter, how to treat risky individual behaviors?

"*Researchers in Spain and Portugal have applied artificial neural networks to the energy grid in effort to predict price and usage fluctuations. […] In short, being able to make adequate predictions based on the patterns of consumption and availability yields far higher efficiency and cost savings.*" How reliable are these predictions? Might there not be a risk of self-fulfilling predictions? How to balance the demand for electricity with production capacity? Will people still be able to set the temperature in their homes?

"*Harvard scientists used deep learning to teach a computer to perform viscoelastic computations, these are the computations used in predictions of earthquakes. […] this application of deep learning improved calculation time by 50,000%. When it comes to earthquake calculation, timing is important and this improvement can be vital in saving life.*" Here too, questions arise about the reliability of the predictions and the eventual automation of warning system. What are the risks of false warnings waking people up in the middle of the night?

"*Companies building these types of driver-assistance services, as well as full-blown self-driving cars like Google's, need to teach a computer how to take over key parts (or all) of driving using digital sensor systems instead of a human's senses. […] These new services could provide unexpected business models for companies.*" Here, we come upon the question about the extreme situations in which the ethical issues are shifted from the user of the vehicle to its designer/programmer. And what about the machine's reliability? How well do digital sensors detect situations? What is the purpose of self-driving vehicles: to transport people while relieving them of the task of driving, or to optimize overall traffic in metropolitan areas?

Let us not overlook the force of the changes that AI might bring about. These changes cannot be separated from the effects of the digital revolution. However a new phase has been entered: certain types of decisions can be conferred on the devices now being developed.[9]

The very design of autonomous, so-called "smart" machines must take account of ethical rules and restraints. These rules are necessary so that the powers delegated by humans to machines will be exercised as we expect. After all, one advantage of AI is that ethical commitments can be traced and verified.

Nevertheless, such measures will have little effect if humans forgo free will. Delegating a power is not an act of abdication. On the contrary, it is necessary to reinforce and develop everyone's aptitude to judge and act with full consciousness and to not let themselves be subjugated by the machines they have built.

---

[8] The four examples correspond to the second, tenth, sixth and first "ways" in GHOSH B. (20 August 2017) "10 amazing ways deep learning will rule the world in 2018 and beyond" available at: http://knowstartup.com/2017/08/10-amazing-ways-deep-learning-will-rule/

[9] COLIN N. & VERDIER H. (2015) *L'Âge de la multitude. Entreprendre et gouverner après la révolution numérique* (Paris: Armand Colin)