Imagine that we are in a perfect world: everything is foreseen, planned in advance, and above all any important decision would be taken rationally on the basis of complete and verified information, never on passing sensations, even less under the influence of emotion. These decisions would cover all important matters, where making mistakes would have serious consequences: medical, hiring, public governance, credit, insurance, or even driving. This perfect, rational world is the one some IT developers boast about.

To achieve this perfection, to avoid costly mistakes, what could be more ideal than to use the latest technological advances, namely Artificial Intelligence? Thanks to neural networks in particular, artificial intelligence has certain capacities for learning, by itself, from the data selected for it, or to which it has been given access. We thus imagine that these networks, having learned “by themselves”, should not suffer from the usual programming errors caused by humans. The machine left on its own seems to make better decisions than any human, or any machine programmed by a human, as illustrated by the case of the self-learning algorithm ALPHAGO Zero, having beaten the previous version ALPHAGO, itself having beaten Lee Sedol, the best human Go player in the world. In short, the more the machine is freed from the human, the better it seems to improve.

Beyond this singular example, where does this representation, this myth, that we carry about technology come from? Why did Karel Čapek's first representation of robots overthrow humanity? Is it an anguish that as creators we inevitably produce monsters that will make us disappear, like in Mary Shelley's novel where Dr Frankenstein was persecuted by the super-human monster he created? More seriously, why do we confusedly think that the machine would become over-powerful, that the Singularity promised by Ray Kurzweil is plausible? To put it another way, why would the machine be less wrong than the human? Why do the results of mathematical calculations seem more rational to us than the faculties of our mind left to itself, than our understanding when it deliberates? According to one of the main representatives of rationalism, René Descartes, adopting the right method would make it possible to avoid making mistakes, except deliberately. After him, it was undoubtedly the consequentialists, in particular the utilitarians, who succeeded in convincing us that error, including moral error, could only come from a miscalculation. Similarly, in a computationalist vision of the mind (where the brain is compared to a computer) thinking is nothing but calculation. If therefore everything, from the knowledge of the true, to the good decision making through moral action, everything, is based on a mastery of mathematics, then, a super computer, a super artificial intelligence software should think more perfectly than our grey cells. As the saying goes, error is human, but fortunately the machine, because of its mathematical perfection, would not be subject to our limitations.

However, recent discoveries of bias in algorithms challenge our beliefs in the perfection of machines. So, should we eliminate error, or does error have a value, whether it is committed by a machine or by a human? We will give two practical illustrations to understand the societal challenges of this question, which is not only philosophical.