Philosophical understanding of AI

Thierry Ménissier

Università degli studi di Trieste
PhD Program Applied data Science and AI
https://adsAI.units.it/
November, 9, 2023, 4 PM Aula Morin
Prof. Thierry Ménissier (Université Grenoble Alpes, France)

Philosophical understanding of AI

1

I'm a professor of philosophy at Grenoble Alpes University in France, specializing in political philosophy and the ethics of innovation, and head of a research Chair called "Ethics&AI".

My academic specialty is therefore what we call practical philosophy, which coordinates ethics, political philosophy, the philosophy of technology and the philosophy of norms.

I'm not a specialist in theoretical philosophy, the other branch of philosophy. But we can't tackle today's issues of AI without posing certain problems with theoretical philosophy. In particular, I'm going to look at the definition of intelligence, and to do this I'm going to talk about some problems in the philosophy of mind and cognition.

One last preparatory point. My definition of AI is as follows: certain mathematical functions, algorithms, enable learning machines to provide information in new proportions and perform tasks previously only achievable by humans. What's more, if AI is not digital, I find it difficult to dissociate them completely. In any case, in the real world, the two go hand in hand.

2

The plan for my talk is as follows: first, I'll introduce this Chair in its context, the Grenoble multidisciplinary AI institute, and then explain its scientific orientation. This will help you understand what I'm going to say next.

Then I'll talk about some theoretical philosophical issues, then I'll go into more detail on the ethics of AI.

In my talk, I'm going to develop a few points that seem important to me – but of course many other important points have come up, which we won't have time to talk about today.

3

I lead a multi-disciplinary research team at a new institute set up 4 years ago in Grenoble.

Grenoble is renowned for the quality of its scientific research (particularly in computer science) and its engineering schools. It is a veritable innovation ecosystem, with universities and major research centers, but also a large number of companies, both large high-tech industrial groups and start-ups.

In this context, the Grenoble site was a candidate for the 2019 call for funding, the aim of which was to bring together French AI forces. This call followed an important report, written by Cédric Villani, famous mathematician, at the request of French President Emmanuel Macron. The title of the report is "giving meaning to AI", and it aims to indicate how France can bring something different to AI. The race for innovation in this sector is dominated by other powers, the United States and China. To succeed in this alternative, we need to concentrate our forces and finance them more effectively. So, once the call for funding was published, there were 12 applications and only 4 proposals were selected. The AI centers funded were one in Paris, Toulouse and Nice, as well as Grenoble.

The first part of the funding lasted 4 years, and the institute is in the process of applying for its renewal for the next six years.

These institutes are based on the work carried out in research chairs. This term designates the activity of a manager who conducts research ty with a team. There are around thirty chairs per institute. In Grenoble, we have a special feature: of the thirty Chairs, 4 are dedicated to the human and social sciences, to support the development of AI. In addition to ethics, these Chairs cover: sociology, legal science and economics. These four Chairs aim both to integrate AI into society, and to provide a basis for legal regulation and ethical assessment.

4

The Ethics&AI Chair isn't just about philosophy. In fact, it is organized around the dialogue between the human and social sciences on emergence and development. During the four years of initial funding, some twenty people have been involved. These include established researchers, as well as post-doctoral and doctoral students.

This Chair does not aim to directly set out the ethical rules applicable to AI. Rather, it focuses on the development of AI in society, on the reception of algorithmic systems by a wide range of social and professional players. It is based on the principle that valid ethical rules do not represent a given reality, but are a horizon. The first step is to observe what is happening with AI in society, according to the different sectors of activity, and then to reflect on the best possible rules for optimizing the development of AI by considering social practices and the uses of algorithms. In this way, it complements the work of computer scientists and roboticists, whose academic work does not focus on these dimensions. In this work, the contribution of various types of knowledge is very important: information and communication sciences, social psychology, marketing science, legal thinking, organizational sociology. I would say that practical philosophy, which aims to think about the best possible rules for AI, brings to this knowledge a questioning capacity to problematize subjects and distinguish the stakes of the situations observed. All this with a view to a desirable AI, in line with the values that democracy sets for itself.

5

All research is based on hypotheses. Those common to all participants in the Chairare the following:

1. “AI” is a common expression, but one that is metaphorical in origin and actually confusing. There is no consensus on its definition, and it still requires, in its various contexts, to be clarified and interpreted
2. The social horizon of expectations with regard to this type of technology appears considerable (in particular, they create trust, which we need to question)
3. These technologies exacerbate the human/machine ambiguities, typical of modernity
4. It is necessary to situate these debates in the context of a philosophy of technology that is in the process
5. The notion of the ethics of AI is itself ill-defined, it covers several different practices and is currently becoming a question of philosophical ethics

6

In practice, our research is divided into 4 programs, which are linked to the projects we have been able to develop.

The program "Methodologies for AI ethics" is a reflection on what the notion of ethics for AI means. It addresses the following themes : Minimalism vs. ethical maximalism, ethics and philosophical ethics, utilitarianism and other forms of ethical reasoning.

The "Socio-technical devices, power and AI" program concerns political philosophy, and questions the following topics: The emergence of new forms of power, surveillance-control-consent, algorithmic society and democracy.

The third program is called : “Ethical issues surrounding the use of AI in healthcare. Emerging care practices, innovative devices, new players and transformation of professions”, and its topics are : Vulnerabilities, autonomy, human-machine dialogue, technologically assisted humans and human enhancement, respect for privacy, redefining health, care and life.

I'd like to emphasize that this program is the one that attracts the most projects. It's clear that the deployment of AI in society can raise very specific and intense ethical issues in the medical sector.

Finally, the "AI, ethics and legal issues" program was added to the other three, as our initial team was joined by researchers in public and private law who could not find answers to the questions they were asking themselves on the following themes: Respect for fundamental human rights, emergence of new concepts / legal personality, redefining the forms of responsibility.

7

The Chair develops numerous projects with non-academic partners. It benefits from the dynamics of Grenoble's innovation ecosystem, as well as from the positioning of academic knowledge that we have chosen: not focused on the academic, but precisely in dialogue with the problems that society is facing with the development of AI.

Here are just a few of the public and private partners with whom we have collaborated, in extremely varied forms, because not all have the same type of questioning, nor the same degree of maturity to accept that the human and social sciences collaborate with computer science and robotics.

When it comes to all these partners, it's fascinating to note that ethical questions don't originate in the minds of philosophers. They arise from social practices themselves, as players in different sectors see their professions and habits profoundly altered by the development of AI systems.

To take a metaphor, today, AI is like a mirror held up to society. The problem we all face is that the image in the mirror should not be so distorted that we no longer recognize who we are.

Our job is first to receive requests and to shape these questions with the help of the human and social sciences, which take the expertise of situations very far. And then to formulate satisfactory ethical protocols with industry stakeholders.

8

This is why, faced with the challenges of innovative transformation imposed by the development of algorithmic systems, the threat comes not from AI, but from the degree of human preparedness or unpreparedness to carry out social, professional and ethical transformations.

This is clear in this lovely illustration recently published in The Times on the occasion of a British government summit.

According to the illustration, the right question with AI is not : *Should we be afraid of deploying algorithms ?*

But rather: *If we humans don't face up to our own moral, social, environmental and political contradictions, shouldn't we rather be afraid of ourselves?*

9

To conclude the first moment of my speech, I'd like to show you what I believe to be the search for a true understanding of AI.

This quest for AI consideration necessarily covers several different dimensions. They can be represented by 4 sides.

The bottom side concerns the definition of AI ethics, and this is where we need to reach a satisfactory conclusion.

But to achieve this, there are 3 other aspects to consider.

Firstly, there is the epistemological dimension. Indeed, as we shall see, it's impossible to say anything about AI without defining the notion of intelligence. We also need to clarify the status of data, since a new relationship to reality (and even a new ontology) is being established with the system of data generated by algorithms.

Secondly, we need to insert our thinking into the real world: that of the economic and social dynamics of innovation.

Thirdly, the right-hand side deals with other concrete aspects, those of defining the good society. I'm talking about the AI-assisted society where people can lead a dignified and free life, and where nature is respected. This dimension concerns social, political and environmental thinking. The first deals with labor issues, the second with the fundamental rights of citizens and good institutions, and the third with the rights of nature to be respected in relation to human activities.

Each side therefore raises questions:

what is the appropriate ethical assessment?

what is intelligence in AI?

Under what conditions can AI as an innovation be genuine social and human progress?

what kind of good society can AI create?

At present, there are no simple answers to these questions. And are perhaps without definitive answers

10

I will now outline a few aspects of the theoretical philosophy of AI.

First point: let's start again with the expression "AI". Is it a misnomer, a metaphor, a fantasy, or hyperbole? Or all of the above?

First of all, it must be stressed that natural intelligence (NI) is not only human: it also applies to animals and plants: in the global sens, natural intelligence is the intelligence of all living species.

To understand how this natural intelligence works, we need to turn to biology (it has long been called "instinct"), and also to the Darwinian theory of evolution. Darwin's contribution is that it took hundreds of thousands of years to produce what we observe in nature, which concerns the relationships that species have with each other and with the world.

Then there's the natural intelligence of the human being: it comes in many forms, and these forms are not reducible to one another. We can distinguish between :

- scientific intelligence, which includes calculation, logic and experimentation
- intelligence typical of arts and crafts: we could call it the intelligence of matter and materials. It's all about finding the right shape for each piece of material.
- finally, aesthetic intelligence, which includes both a sense of nuance and a sense of composition to create balanced ensembles.

There are, of course, similarities: for example, the first and third forms mobilize intuition, which is important in both science and art.

Based on these quick characterizations of natural human intelligence, can we establish that machines do (or don't) think?

When considering the intelligence of machines, we need to evoke a famous argument, the Turing test (1950)... but this thought experiment is debated, as you may know.

11

Alan Turing invented the Turing test as part of his famous article entitled "Computing Machinery and Intelligence". This paper was the first attempt to define intelligence and establish a way of measuring it in a machine. The test was based on a game called "Imitation", in which a man and a woman tried to convince a third player that they were of the opposite sex.

Alan Turing proposed that if a machine could convince a person that he or she was a human being in the course of a conversation, then that machine could be considered intelligent.

How does the Turing test work?

A person speaks to a machine via a keyboard and screen. The person doesn't know whether he or she is talking to a machine or a human being. If the person can't tell the difference, the machine has passed the test.

Most variants of the test take place in three stages.

In the first stage, a human being talks to a machine and another human being, both without knowing which is which. The aim is to establish a natural conversation, as if interacting with another person. The conversation takes place via a terminal.

In the second stage, the examiner asks the test questions and expects the answers to be convincing and plausible. The aim is to determine which of the two interlocutors is a human and which is a machine. If the examiner is unable to distinguish between the two on the basis of the answers, the machine is deemed to have passed the test.

The final step is to evaluate the results. Passing the test means that the machine has made the examiner believe it was talking to another human being. In this sense, it is considered to have demonstrated intelligent behavior similar to that of a human being. Logically, however, this test is not considered a valid indicator of artificial intelligence. This is because the test is based on a kind of bluff, which consists in making people believe in intelligent behavior. The Turing test is a controversial method for evaluating AI. While some consider it a useful tool for measuring

progress in this field, others claim that it is flawed and does not assess true intelligence. Despite this, the test represents a historic milestone in the development of this technology.

12

A recent contribution explains why this argument has been important for the whole direction of AI history. It's by Daniel Andler, specialist in mathematics & philosophy of mind and cognition, and his book on the double enigma.

Andler considers that AI poses the problems of knowledge in the same way as the philosophy of mind did before it appeared: it pursues the questioning in a different way.

For him, AI is an enigma, as natural intelligence always is.

An enigma is neither a mystery (incomprehensible, mystical) nor a problem (that can be solved rationally).

13

Why, for Andler, *is AI an enigma?* For two different reasons

1. It violates the "verum-factum" principle:

This is the principle that the best way to understand something is to make it

This is the principle called Convertibility of truth and fact, proposed by the Neapolitan philosopher Giambattista Vico in "La Scienza nuova". (Vico, 1744)

But, continues Andler, even if we calculate the algorithms, they produce results that we can't totally anticipate. AI" results is unpredictable.

Consequently, computer science must be both formal and empirical

2. Andler notes that the closer AI tries to approach the performance of natural intelligence, the less it resembles it. It is therefore something else, something new and creative, something we don't yet know very well.

Andler's conclusion is : computer scientists will never be able to recreate natural human intelligence.

14

Andler goes on to provide an original definition of human intelligence: it is a normative relationship to the concrete situation of an human animal.

It is not the ability to solve problems or perform calculations

Problem-solving and calculation are not the primary functions of the IHN. We don't know what its primary function is.

Which is perhaps not just an enigma, but a real mystery (my opinion, TM).

Andler therefore proposes a pragmatic (in the philosophical sense of the word) approach to human intelligence, not a functionalist definition.

For his part, he believes that AI is above all a means of solving problems: in his case, a functionalist approach is necessary.

Second consequence: the difference in nature between the two kinds of intelligence means that computer scientists should not try to reinvent natural intelligence.

15

The problem is that the desire to recreate the human intelligence is at the origin of the AI project.

This can be seen by reading the last paragraph of Turing's famous article.

In this paragraph, he imagines what learning machines will be like.

You can read what Turing wrote.

For Turing, when he invented the AI problem, there were two possible ways of making machines intelligent. He says he is unable to say which is the best path. The first (1) consists in enabling them to be the best possible from the point of view of pure calculation. The second (2) is to equip them with systems that provide them with information: on the one hand, the best possible sensory organs (2.1) and, on the other, the language that enables us to name things (2.2). However, it is this second option that has created ambiguity, and enabled the dream of creating another Human and Natural Intelligence with AI.

16

The debate concerns the isomorphism between human and machine.

The question was whether the equivalence between machine and human should be high or low: this distinction divided AI pioneers. The proponents of high equivalence were Herbert Simon and Allen Newell, who argued that natural human intelligence feeds on information that enables it to perform functions that a machine cannot. For them, AI is "anthropic".

The proponents of low equivalence are Marvin Minsky and John McCarthy. For them, AI must be freed from the obligation to resemble natural human intelligence. What matters is that it achieves certain results, not that it does the same thing as natural human intelligence. Consequently, the AI they have imagined can be called "ananthropic".

17

This springtime of AI could be described as a "Promethean dream", in reference to the titan of Greek mythology who created man, or emancipated him from the gods by giving him fire.

But this dream failed, and, in Andler's opinion, came the first winter of AI.

The upside is that AI projects have abandoned their profane dimension (for it was a kind of profanity to try to match natural human intelligence with artifice). The techniques used have become industrialized, and AI has established itself as a profession for engineers, capable of building reliable machines.

18

This debate now seems outdated in the scientific community: AI has emancipated itself from the model of human intelligence. This is why it has become a self-confident technology. And as a technology, it possesses three major characteristics that make it exceptional: it is indeed :

1. Universal: it establishes a continuum between science, technical objects, uses and social practices,

2. Structural: it penetrates and links all spheres of human activity.

3. Possibly unlimited: it is capable of overturning, transforming and renewing all areas of human activity.

With such formidable characteristics, what can we compare it to? Electricity, which appeared three centuries ago? To written language, which appeared much longer ago? I think it's a fascinating question. Perhaps nothing can compare with AI!

19

Nevertheless, AI stimulates the imagination and continues to fuel popular fantasies. Here we find the overconfidence with which this formidable technology is viewed.

In this regard, my second point concerns the French baptism of the word computer.

I'll tell you a story about this, which took place at the same time as the American debate on isomorphism. You may know that the French word for "computer" is "ordinateur". Why this strange translation?

20

This man is responsible : Jacques Perret, full professor of grammar at the Sorbonne University in Paris and Catholic theologian. He is the author of several books, including : *A study of the idea of Rome. The legend of Rome's origins ; Virgil: man and work ; The Accomplishments of Eternal Life*.

This was a time of very large machines, such as the ENIAC at the University of Pennsylvania. These large machines are an interesting contextual detail for Perret's idea.

21

Professor Perret was consulted by the President of IBM France to find a term to translate "computer".

And you can read his reply.

“Dear Sir,

How about "ordinateur"? It's a well-formed word, even found in the Littré (the French reference dictionary), as an adjective for God who puts order into the world.. »

It's an extraordinary response !

22

To be honnest, we should add that in his letter Perret also proposes other terms:

Combinateur, which would be a kind of mixing machine.

Synthesizer, and Sélecto-systémateur is a functionalist description, since it refers to two basic operations: selecting information and synthesizing it.

Congesteur and Digester : these last two terms are based on a bodily or organic metaphor: congestion, in physiology, is the influx of blood into a part of the body; digestion is the way in which food, by breaking down, nourishes the body.

and even, with a feminine formula: « electronical Ordinatrice » which is untranslatable in English. And this wording perhaps evokes memories of the six women programmers who operated the ENIAC, sometimes call the "ENIAC girls".

But at least, with Perret's semantic innovation, we know what a computer looks like (at least for someone who speaks French).

Here's what it looks like: it's something divine (in this case, not exactly God, but the pantocrator Christ of the Orthodox religious tradition).

It's interesting, even disturbing, that the inventor of cybernetics, Norbert Wiener, also drew a parallel between computer science and theology in a strange text. Perhaps computer science, with AI, is indeed attempting a theological ordering of the world!

23

In this second part, I'd like to ask what we mean by AI ethics.

Following on from the first part, the general problem is formulated as follows: How do we deal with a non-human, but God-like super-intelligence?

(at least, that's what Perret's translation would have the French believe).

As a preamble, I'd like to make 3 comments.

First comment: AI ethics is a work in progress... Work is just beginning.

there are already many global declarations and charters (Jobin, Ienca & Vayena 2019) but with few common defined concepts...

- This raises two unanswered questions:

- At what level of the "AI phenomenon" does evaluation & regulation

- Who will set it out & apply it?

24

The second comment concerns the comparison between AI ethics & bioethics.

Experts often compare the two attempts.

- At first view, the two approaches share similar principles:

For bioethics : Beneficence, non-maleficence, autonomy, justice (Beauchamps & Chilldress, 1979: Principles of Biomedical Ethics)

For artificial ethics : Respect for human autonomy, harm avoidance, fairness, explicability (High Level Expert Group on AI appointed by the European Commission, 2019)

- However, the comparison is limited:

- we can't compare the ethics of care and health / the evaluation of a technical system: in fact, there are very different expectations and socio-economic realities, as recently demonstrated by other experts of the French Comité national pilote d'éthique du numérique, 2022

25

I'd like to formulate my third comment directly with a question: What are the ethical risks posed by AI?

This question touches on the interest or relevance of the ethical approach.

Or, It's not clear!

For engineers, the risk is when the technology is badly designed. It's not ethical but certification problem.

However, this notion of risk has recently become very important for AI technologies as the European Union would like them to be designed. The future IA Act regulation even mentions an ordered or hierarchical typology of risks: unacceptable, high, limited, low and minimal.

And experts have begun to identify technologies according to risk levels.

So we can see that the notion of risk exists outside the purely scientific and technical dimension, at least for European regulation. We can also see that these risks concern practices and uses rather than the technologies themselves.

But the ethical risks are another matter.

So what is an ethical risk when you're a scientist or engineer?

26

To understand this, we can start from the opposite of the ethical attitude, by saying :

To act without ethics is without reflecting on one's own actions.

What does it mean?

To shed some light on these terms, I turn to a great philosopher of totalitarianism, Hannah Arendt. In one of her most famous works, Arendt recounts the trial of Adolf Eichmann, who was tried as a war criminal. The title of her book includes the concept I want to talk about: the banality of evil.

Eichman was not a Nazi with a military vocation, like Goering or Goebbels. He was merely a high-ranking German civil servant serving the German Reich. His job was to be the chief rail regulator: he had to organize german rail transport. In his defense in court, he claid not to know what was on the trains for which he had superior responsibility. On those trains, of course, were human beings (Jews, Communists, homosexuals, political opponents) deported to concentration camps.

How far, in a totalitarian state, is a zealous official like Eichmann responsible for the consequences of his actions? Of course, civil servants cannot be held responsible for all the consequences of their actions in the performance of their duties. But can we totally ignore them without abandoning our own moral conscience?

Arendt takes into account the arguments of Eichmann's defense, reflects on them philosophically, and explains this: in a totalitarian state, all humans can be like Eichmann. The banality of evil consists in accepting evil without concern. It's the same as disregarding the consequences of one's actions, and being uncritical.

This argument of the banality of evil is very powerful because it stimulates moral conscience, which begins with the scruple of knowing the consequences of acts.

It's interesting and significant that it has reappeared in the recent AI debate: a few months ago, it was used in the *New York Times* by the famous linguist Chomsky (with two co-authors) in a critical article written against the use of generative AI: their "false promise" covers a new form

of banality of evil. The ethical risk," writes Chomsky, "is the rise of indifference and the loss of a critical sense ».

27

A similar unethical attitude can be observed in the field of science and engineering, with an example well known to historians. Fritz Haber, the german chemist who discovered the process of synthesizing ammonia, was awarded the Nobel Prize for Chemistry. His wife Clara Immerwahr was also his colleague. And when he was tempted to use her process to make military poisonous gases, she told him that for her, this temptation betrayed the meaning of science. Haber didn't listen to her and left to supervise artillery fire at the front, with gas-laden shells. Immerwahr refused to accept this situation, as she found it morally unbearable, and committed suicide, as she had threatened to do.

Haber's behavior raises the following two questions:

Is Haber a rigorous scientist, a German patriot or a mass murderer?

and on another level : What is the purpose of science?

And with this double example, so we have two reference positions: on the one hand, indifference to the consequences of scientific and technical discoveries (Haber), on the other, hyper-responsibility for the consequences (Immerwahr).

Questioning the risks of the consequences of discoveries and inventions, that's the level of understanding of ethical risks. It's always a question of positioning yourself in relation to these two options.

28

Of course, these considerations may seem rather gloomy, even inappropriate. But in fact, it's easy to see how appropriate they are, once you restore them to their popular formulation, so famous and so obvious to common morality.

What is this formulation?

It's the Peter Parker principle.

This principle has a very long tradition in moral philosophy, and even seems universal. Stan Lee recently reformulated it, giving it the force of Peter's uncle Ben Parker. The initiatory journey of the teenager who becomes Spider-Man is to succeed in living according to this principle. For a teenager, this is no easy task. The superpowers of today's AI put computer scientists in the same position. What do you do when you suddenly have so much power?

29

Now I return to our main question: What ethics for AI?

I start from a research hypothesis: it is possible to Identify, distinguish & combine the specific or "regional" ethics of AI

30

I have explored this hypothesis in my research. And I identified four different ways of practicing ethics for AI, digital and data. This is what I call the 4 AI ethics. These are different, specific discourses, with different methodologies. The 4 ethics are, in the words of philosopher Michel

Foucault (*The archaeology of knowledge*, 1969), ethical discursive formations applied to AI and digital technologies.

Computer ethics concerns the design of AI systems: it is practiced by computer scientists.

Artificial/algorithmic/robotic ethics concerns the design, programming & uses of autonomous intelligent machines: it is practised by roboticists and users.

Digital & data ethics concerns the design and use of platforms in the innovation dynamic of the digital economy: it is practiced by economic players, users, and aims at Creation and data management, where users, public authorities and economic players act.

Finally, what I propose to call UX AI ethics (UX: referring to user experience in innovation), concerns the selationships between AIS and their meaning and values in society. It involves social scientists and philosophers, users/citizens and public authorities. And it's based on co-design, participation and stakeholders' involvement.

Importantly, the first form is dominant, which is normal, since it's about examining what creates algorithms.

But the problem is that it implicitly gives primacy to a single form of ethical reasoning (the so-called consequentialist form, with utilitarian ethics). Our ambition, in the Ethics&IA Chair, is to mobilize other forms, for better and more complete ethical assessments.

31

Let's start with computer ethics.

This form of AI ethics concern the work on the values that should govern code writing.

These values have been clearly identified and are the subject of a large body of computer science literature. For example, the article by Barredo Arietta et alii. represents a very good synthesis of them and indicates how work on these values constitutes a new branch of computer science research, named Explainable AI. This new field aims to clarify the notions deployed in computer ethics with a view to accounting for and understanding the approach taken when designing algorithms.

We can distinguish between epistemic values and social values.

Epistemic values attest to the ability of algorithm designers to master their creations. The first value is explicability. In a way, it represents the basic imperative of a deontological ethic of computer science; indeed, it is fundamental because it is thanks to it that the work can present the guarantees of science, whose discourse is always demonstrable.

Among social values, several have meaning not only in relation to algorithms, but far beyond. For example, for computer ethics, fairness means ensuring that there is as little bias as possible; but for moral and political philosophy, it is of course an extremely important theme (since antiquity, the first complex theory of fairness dates back to Aristotle's *Nicomachean Ethics*).

The other important subject is to specify how, in addition to helping to write code, these values can be active. For several years now, we've been seeing a lot of declarations in favor of such values, from those who build computing machines of course (like IBM) or companies who trade in the digital network (Meta), but also from companies in all sectors of activity (for example, in the insurance sector, Generali). Everyone seems to be committed to the same values, but of course this is only declarative.

I would like to point out that another possibility has emerged, notably with the collective experience of the Montreal Declaration for the Development of Responsible AI (2018). This experience is different and remarkable because, in this case, there was a truly varied stakeholder participation: computer engineers, social scientists, representatives of the population, to elaborate not only technical, but social principles.

32

With regard to the activation of values, it concerns the guarantee that the values of computer ethics will be applied. In addition to organizational charters, another possible approach is an oath for computer scientists, comparable to that taken by medical doctors. This is the Hippocratic oath.

It is always linked to national medical associations (« Ordre des médecins » in France, General Medical Council (GMC) in UK, American College of Physicians or American Medical Association in US.) which regulate and control the practice of the medical profession. These professional associations can impose constraints such as exclusion, making it impossible to practice medicine.

It's worth noting that since 2018 an attempt has been made to establish an oath for computer scientists, the so-called Holberton-Turing oath, with double reference to Allan Turing and Margaret Holberton, one of the computer scientists who programmed ENIAC at the University of Pennsylvania.

33

It's worth investigating this further: can an oath like the Hippocratic Oath complement computer ethics?

The old version of the Hippocratic Oath, which is neither abridged nor simplified, shows several things. I'm going to make just 2 comments:

1. Note the force of the expressions used: the appeal to the gods, the willingness to commit oneself before the human community, the risk of completely losing one's reputation. This point is probably different between doctors and computer scientists, because with the former, the life and death of patients and patients is at stake! He therefore points out that the Vocation to care merges with medical ethics, and is part of a personal moral commitment to doing good for humanity. This point is also different, because while a computer scientist may want to commit to doing good, this is not directly what is expected of him or her. Doctors are necessarily more willing to apply ethical principles than computer scientists.

2. At the same time, through its appeal to social reputation, the Hippocratic oath demonstrates that medical deontology lies between social constraint and moral vocation. Not to mention the existence of professional associations to ensure that medical values are applied in practice. This point reveals that the force of the oath is not only moral (if by that we mean something that comes from the individual taking the oath), it's something social. And this is compatible with the profession of computer scientist.

34

This is all the more true given the long-standing existence of oaths for engineers around the world. For example, the Rite of commitment for Canadian engineers (since 1922). It gives access to the symbol of possession of a ring. Like the magic ring in The Lord of the Rings, except that the first hundred were forged from the steel of a bridge that had collapsed, causing

deaths. So, with this solemn declaration, we are committing ourselves to guaranteeing the safety of all users of the artifacts built by our engineers.

It should be noted that :

1. Commitment and possession of the ring is not compulsory, but purely voluntary.

2. it is not a prerequisite for entry to the engineering profession in Canada

3. The Corporation of the Seven Wardens Inc. is the independent association that is responsible for administering the engineer's rites of engagement.

We can see that this corresponds to a superior professional certification. Here again, the social and moral dimensions frame the engineer's vocation.

But it's also worth noting the power of the words the individual utters during the ritual.

35

That's why it would be interesting to propose A possible workshop exercise that allows you to understand the power of oaths by experiencing it directly. Even if it's not a real oath, it's very interesting to experience it personally in the form of a game.

This workshop can be used as part of an engineering school curriculum. I do it myself in my courses with engineering students. It should include the following moments:

1. Specify several important ethical values to your profession, e.g. honesty or integrity, the desire to create beautiful and useful things, etc.

2. Order them by priority: what is the most important ethical value for practicing your profession with dignity?

3. Draw up an oath for the engineers in your workgroup

At this stage of the drafting process, ask yourself on which higher power specific to your profession the oath is founded.

4. Declare this oath in the presence of your family and professional community. In this stage, the declaration must be solemn, so you need to mobilize certain emotions such as gravity and commitment, to make it seem as if it were a real moment of oath-taking.

After these steps, you have to ask yourself: how does this affect your activity?

Listening to the engineering students who played the game, I realized that they were expressing what the philosopher and linguist Austin called the performativity of language (he was talking about « speech acts »). I conclude that commitment to a profession with solemnly declared values represents a guarantee that the values of computer ethics can be effective.

36

And now I come to UX AI ethics, which we have combined in our work with codesign experiments.

Based on AI uses identified by the stakeholders themselves in various social practices

It aims To involve all stakeholders:

AI designers, prescribers, users & social scientists

(economy, social psychology, anthropology, political science, philosophy)

The design and evaluation of AI systems go hand in hand.

And to do this, we need to express the various points of view, the explicit contradictions in values, principles, interests & forms of reasoning.

We need to highlight dissensions & clarify choices.

From the outset, two forms of ethics are mobilized: that of discussion, proposed by Karl-Otto Appel and Jürgen Habermas, and that of capabilities, proposed by Martha Nussbaum.

We can see here too (as with the theory of oaths) that there are many tools of moral theory to accompany AI, digital technology and data on the ethical way.

37

I'd like to emphasize a point very important to me. This way of practicing ethics based on usage mobilizes various stakeholders. It's an opportunity to go beyond (naive) utilitarianism & go against moral minimalism, which is all too common, using the 4 forms of ethical reasoning

The protocols we have set up in Grenoble with our partners enable us to train participants in applied moral philosophy while dealing with AI projects.

The forms of reasoning mobilized concern what we call deontologism, axiologism and aretism.

Deontologism means using one's conscience to find the rules one wants to follow because they are good. A deontological answer in AI would be to ask whether such and such a system of algorithms does not violate what is sacred to a moral conscience.

Axiologism means appealing to a higher value, as in the case of ecological ethics: nature, life, GAIA (Latour). In this case, an axiological response in AI or digital technology would be to ask whether the same system is not ruining the planet's resources.

Aretism is about becoming a better person. In this case, the aretism answer in AI or digital rests on the question of whether the system makes people better or worse.

38

The question here is whether a non-consequentialist ethical approach to technology is possible.

There are in fact two questions:

- is it possible?

- is it relevant?

The answer to the first question is yes. This makes it possible to To resist the possible harmful effects of technologies & to support their development for better practices

It is possible to reason ethically about AI with other forms of reasoning. Recently, for example, colleague Shanon Vallor of Edinburgh University in Scotland proposed a remarkable analysis of virtue ethics applied to technology.

The answer to the second question is: maybe.

Indeed, this perspective has a high degree of ambition and involves a great deal of work.

We could say that it pursues two objectives, one minimal and the other maximal.

1. The minimalist objective is Ensuring that technological uses do not prevent us from leading a moral life. Indeed, this is sometimes the case, or may be the case: some AI technologies can make us worse, and prevent us from leading a moral life.

2. The maximalist objective can be formulated as follows: Becoming ethically better through technology. Personally, I find that it meets the vocation of computer scientists who want to create quality tools for a better life. But utilitarianism in a purely economic sense often fails to achieve this, and technology is used solely for material profit, whereas it can contribute to human dignity and happiness.

39

Concluding remarks : 3 final comments.

Firstly, ethics are by no means irrelevant to work on algorithms. It's not an extra, but, given the widespread deployment of AI in society, it's a major social necessity.

Secondly, ethics isn't just about risk reduction. It's about the world that's desirable. Given the variety of viewpoints that make up democratic societies, what computer scientists are discovering today is that there must always be room for collective discussion of values. Debiasing algorithms is necessary, but not sufficient. If this work is not to be naïve technosolutionism, we need to be trAIned in moral theories. This can be done by taking part in collective co-design workshops, before returning to work on algorithms.

And finally, I believe that AI and digital technology offer a real opportunity to improve the world, while remaining human with super-human technology!

Thanks for your attention !