



HAL
open science

Decoding AI in 2025

Valérie Sagant, Harold Épineuse

► **To cite this version:**

Valérie Sagant, Harold Épineuse. Decoding AI in 2025. Decoding AI in 2025, Jan 2025, Paris, France. pp.70, 2025, Actes. ⟨hal-05497970⟩

HAL Id: hal-05497970

<https://hal.science/hal-05497970v1>

Submitted on 29 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License



DECODING AI IN 2025

INSIGHTS INTO A TECHNOLOGY
ON ITS WAY OF NORMALIZATION

Conference

JANUARY 23, 2025

The **Actes** collection offers written proceedings of various events (conferences, seminars, workshops, etc.) organized by the Robert Badinter Institute, formerly the Institute for Studies and Research on Law and Justice.

It records the enriching exchanges between speakers and participants on topics that bring together perspectives from both the professional and research communities.

Each publication in the collection is an influential document for the sharing and valorization of practical and scientific knowledge on law and justice.

3	Foreword
5	Welcome address by the organizers
8	But where is AI going?
15	Speaking the language of generative AI
20	Will AI transform freedom into destiny?
36	AI and legal professionals, a great misunderstanding?
57	AI, a weapon of mass destruction?
63	References
64	Conference program

Foreword

The “Decoding AI in 2025” conference brought together lawyers, judges, academics and artificial intelligence researchers, to shed light on these complex issues through a multidisciplinary approach. Discussions confirmed the benefits of a considered and informed approach to AI, emphasizing ethics, responsibility and the need for training. As AI continues to evolve, the role of lawyers and legal institutions will also need to adapt to ensure that this technology is used for the common good.

The “Decoding AI in 2025” conference, organized by the “Institut des études et de la recherche sur le droit et la justice” (IERDJ) and the University of Strasbourg’s Faculty of Law, Political Science and Management, in collaboration with the “Centre de culture numérique” of Strasbourg’s University, was an opportunity to set the scene for a much-needed reflection on the institutional and professional uses of AI in the field of law and justice, through a reminder of a few popularized technical basics, a review of the political and societal stakes of AI at the start of 2025, and an updated, concrete and practical state of the use of AI in legal and judicial professions.

The conference opened with an address by the Dean of the Faculty of Law, Political Science and Management, who expressed the growing importance of AI in various fields, including law. She emphasized that AI is not simply a technological phenomenon, but a revolution that is influencing our way of life, our social relations, and even our conception of human identity. The need to “decode” AI has thus become significant to understanding its ethical and social implications.

The challenges of artificial intelligence for lawyers

The speakers focused on the challenges and opportunities presented by the introduction of AI into the professional practices of lawyers. They highlighted the need for an in-depth understanding of the underlying technologies and algorithms, to better grasp their impact on the legal system.

An important point discussed was the regulatory framework surrounding AI developments, with a particular focus on the European AI Regulation. Speakers expressed concerns about vague definitions and the difficulty of applying legal standards to constantly evolving technologies. The importance of training and education was also highlighted, as it is essential that legal professionals are equipped to interact with these new technologies.

The applications of artificial intelligence in the legal field

During the conference, concrete use cases of AI in the legal field were presented. Lawyers and magistrates shared their experiences of integrating AI into their practices. For instance, some have started using AI tools for legal research, case management, and even document drafting. However, they also highlighted the limitations and risks associated with these tools, particularly in terms of liability and reliability.

The use of generative AI in the judicial process was also addressed. Speakers discussed the ethical implications of using algorithms to influence judicial decisions. The importance of transparency and accountability was highlighted, with the idea that judges must remain vigilant about the tools they use.

The evolution prospects

Participants discussed emerging trends and potential developments in the field of AI. The need for interdisciplinary collaboration between lawyers, technicians and researchers was highlighted as a key factor in navigating this complex landscape. The importance of being informed about technological developments was underlined, as was the need to take a proactive approach to integrating AI into the legal field in an ethical and responsible manner.

Speakers also expressed concerns about the impact of AI on employment in the legal sector. While some tasks can be automated, it is crucial to preserve the human role in the judicial process. Particular attention was paid to training future lawyers to work efficiently alongside AI systems.



Further information

Audiovisual recordings of the various sequences and PowerPoint presented during the conference provide additional support for a deeper understanding of these crucial issues.

Audiovisual recordings and support materials are available online following this link:

<http://bit.ly/3JswtyJ>

Welcome address by the organizers

Jeanne-Marie TUFFERY-ANDRIEU

Dean of the University of Strasbourg's Faculty of Law, Political Science and Management

Yannick MENECEUR

Magistrate and associate lecturer at the University of Strasbourg's Faculty of Law, Political Science and Management

Harold ÉPINEUSE

Deputy director of the "Institut des études et de la recherche sur le droit et la justice" (IERDJ)

Jeanne-Marie TUFFERY-ANDRIEU: Dear deputy director of the IERDJ, Strasbourg's judicial tribunal president, Colmar's judicial tribunal president and Deputy Vice-President in charge of training. The Faculty of Law, Political Science and Management is delighted to be taking part in this wonderful event, entitled "Decoding AI in 2025, insights into a technology on its way of normalization", at the invitation of Mr. Yannick MENECEUR, a magistrate specializing in digital issues and associate lecturer at the faculty. I would like to express my deepest gratitude to him, to which I am honored to associate the IERDJ, particularly Valérie SAGANT and Harold ÉPINEUSE.

I am also delighted to welcome colleagues from other universities, and to salute the expert practitioners who have made themselves available today. This event, and the presence of the students enrolled in the courses associated with the law specialization, reveal the importance for the faculty of the issues raised by the emergence of AI and digital technology in the fields of law, justice and numbers. Far from wishing to follow a fashionable phenomenon, the faculty is determined to put down roots in this dimension, with which it is already familiar. The training courses it has been offering for almost 30 years now—and

here I would like to pay tribute to the ongoing commitment of Ms Catherine Ledig, Associate Professor at the Faculty in these fields—underline the importance that digital and technology have always had at the faculty. The presence of the Deputy Vice President also reflects the university's special attention to these issues. Mr. Deputy Vice-President, in the wake of the day devoted to the use of artificial intelligence that you organized, seeing you here supports our project and encourages our ambition. It is clear that, without specifically earmarked funding, without additional resources, our determination alone, no matter how determined, will achieve nothing. We are therefore counting on you, Mr. Deputy Vice President, and through you on the university, to enable the faculty to remain active, dynamic and innovative in this field, but also to enable it to contribute to the decoding of AI thanks to the values it has brought since the 16th century, those of the reigning humanism which is its cradle and which still leads it to assert the pre-eminence of civil liberties and respect for the dignity of the human person.

Supported by the academic and professional experts who drive our training and research, the faculty is aware of its responsibility to the city. Artificial intelligence is an extremely powerful tool used in many areas of human activity, and we can now assume that its use will increasingly influence our way of life, our social relationships and even, in the future, the way we conceive our identity as human beings. To want to decode AI is also to recognize that the advent of artificial intelligence represents a genuine revolution, announcing the creation of a new social system. But wanting to decode also means asserting the will to discover the coding key to make it more accessible.

Does asking this question means that the tool has escaped its creator? By extension, does this mean that we are witnessing the disappearance, or at least the eclipse, of the meaning of the human being? Would this lead artificial intelligence programs to call into question the human

being and his actions? Decoding AI therefore also means raising the question of the ethical dimension of decision-making, a question that considers not only the results of an action, but also the values at stake and the duties that flow from them. I would like to take this opportunity to express the faculty's pride in being involved in this reflection, which contributes to the scientific and human training of students, to the implementation of a truly social policy and to the development of a just city. I wish you very fruitful exchanges.

Yannick MENECEUR: It is with great pleasure and honor that I find myself here in this amphitheater, in your presence, after several years of reflection on the most appropriate way to share my thoughts on this technology. I would like to express my deep gratitude to the Dean for her confidence in me, and for allowing me to take up the position of Associate lecturer at the Faculty of Law. It has been a real struggle, but we have achieved it.

I would also like to pay tribute to the presence of the Deputy Vice-President, who showed great sensitivity. Our schedules did not quite line up, but we were able to reorganize our days accordingly, and I thank him for his attention.

I would like to greet my colleagues in the room, as well as those who are online. I made myself known within the network of the "Ecole Nationale de la Magistrature" (ENM) and informed my colleagues of the existence of this conference. Nor should I forget the IERDJ, in particular, Valérie SAGANT and Harold ÉPINEUSE, who are at my side and to whom I shall soon hand over the floor. Finally, I would like to greet my students; I am taking ownership of your presence a little bit, even if other Professors are also here to look after you.

I have organized this conference for you. Today, it is mandatory to popularize artificial intelligence. Article 4 of the AI Act imposes specific training and knowledge requirements in the field of artificial intelligence for all professionals. So, whether you are in training or already practicing, it is essential that we master at least the basics of this technology. Just as we distinguish between diesel and petrol in a motor vehicle to choose the best technological solution, we also need to understand the essence of this technology.

We are not pretending to turn you into computer engineers or specialists, with respect due to the artificial intelligence researchers present

today, who will build bridges between our knowledge. We want to go beyond the prefabricated discourses and often biased narratives surrounding artificial intelligence and bring this technology in the real world. This day is all about pursuing these goals.

Harold ÉPINEUSE: It is a pleasure and an honor to be here again, as part of a new collaboration between IERDJ and the University of Strasbourg's Faculty of Law, Political Science and Management. I would like to express my gratitude to the Dean for her confidence, and to you, dear Yannick, and Catherine Ledig, for your unfailing support.

I would also like to thank the "Centre de Culture Numérique" for welcoming us to its premises today, providing us with all the resources we need to organize this event here, both onsite and online, to reach a wider audience. This public conference marks the launch of a series of closed workshops with our institutional and professional members, during which we will reflect together on the issues surrounding artificial intelligence in the field of law and justice.

Who are our members? They come from the Ministry of Justice, the judicial and administrative jurisdictions, and the various legal and judicial professions. We also include the Ministry of Higher Education and Research, the CNRS and France Université, among others.

By mentioning this, I wish to illustrate that our Institute is positioned at the crossroads of institutions, the legal and justice professions, and the world of research. The IERDJ is, in a way, the common home of researchers and professionals in the legal and justice fields, with the aim of supporting studies in these areas, as well as promoting and funding multi-disciplinary research. I would like to emphasize that we are not going beyond the law scope. With the diversity of members I have just mentioned, we already bring together some of the best legal minds.

In that respect, we bring together lawyers, institutions and professions among the field of humanities and social sciences, to better understand the phenomena affecting law, justice, our institutions and the evolution of our professions. Digital technology has been at the heart of our work for almost fifteen years.

We are not alone in this approach. The question of AI, which is our focus today, is a theme we have been exploring at the Institute since 2016,

in collaboration with the Council of Europe and the University. At the end of our work on cyber-justice, we organized a conference during which we asked a question that seemed iconoclastic at the time: is artificial intelligence an enemy or an ally of the law?



Conférence Cyberjustice Europe 2016 - Intelligence artificielle : alliée ou ennemie du droit ?

<http://bit.ly/3LbOFFs>



Asking this question today may seem less relevant, or it may take on renewed meaning in the light of recent developments in artificial intelligence. Many of the questions raised at the time were linked to the doctrine and policy of open data, which was beginning to take root in the field of law and justice. This generated debates around the notion of predictive justice, a field of reflection that you have documented very well, Yannick.

Recently, about two years ago, a mass consumption tool emerged, namely generative models, which gave new impetus and new perspectives, requiring a renewed look at the role of artificial intelligence. We are seeing just how embedded AI has become in our personal and professional lives.

As you can see from the program we have given you for today, we are not going to deal straight away with the impact of artificial intelligence on the work of lawyers, in the broad sense. Instead, we will be taking time this morning to explore AI as a whole and its recent developments, considering the current state of AI in January 2025, rather than 2016, as I mentioned earlier. Everything is evolving very rapidly.

We are flooded with information about AI. This conference comes at a time when many events

are being organized on this subject. I hope this one will add real value, and you can let us know what you think at the end of the day. However, it is important to recognize that we are overloaded with contradictory information and news stories, and we have recently heard of massive investments on the other side of the Atlantic to develop an even more powerful and widespread AI, which will impact our lives.

Before we start the workshops with our institutional and professional members, in which we will examine the frameworks for AI use in our businesses, as well as the opportunities and limits it presents, it is essential to start by reflecting on the latest developments and thoughts sparked by this technology from the experts and researchers we have invited this morning and afternoon. I hope we will have the opportunity to engage in a succession of views and a multidisciplinary dialogue. Thank you all for coming. There are many of you here in this amphitheater of the "Centre de culture numérique", as well as many of you online. I think it is time to hand over the microphone to our keynote speaker for a long-awaited inaugural lecture. ■

Opening conference: But where is AI going?

- Understanding the foundations and ethical implications of artificial intelligence
- Taking a step back from preconceived ideas about artificial intelligence
- A critical reflection on the motivations of major technological players is in order



Jean-Gabriel GANASCIA

Professor of Computer Science at Sorbonne University's Faculty of Science, senior member of the "Institut Universitaire de France" and Chairman of the CNRS Ethics Committee

Firstly, I would like to thank you for inviting me to give this inaugural talk in front of a group of lawyers, which is admittedly somewhat intimidating. The aim, of course, is to talk about where AI is headed. However, to know where AI is headed, we first need to understand where AI comes from. This is the subject I would like to address, without further ado, in this opening lecture, by introducing the key concepts present in this scientific discipline.

I have been asked to give a general presentation. So, let us get to the heart of the matter. The aim of my book, as already mentioned, is to give the keys to artificial intelligence to a wide audience, both scholars and teachers. Indeed, in my opinion, this should be part of general knowledge. Unfortunately, when you listen to the media, even though there is a lot of talk about artificial intelligence, I feel that it only adds to the confusion. Since the subject is so hotly debated,

people from very different backgrounds are invited to talk about it, and sometimes, some of them are not experts or only have practical experience of artificial intelligence, or even just general ideas, without really knowing anything about it.

This short book presents a scenario, that of four young scholars who come to see me to write a special issue of their school's journal—the middle school Raymond Queneau—entitled "Querelles de Queneau", at the request of their French teacher. I am not going to discuss the content of the book itself here; I simply want to mention the plot, which is, of course, completely imaginary. These middle school students were invented, as was the French teacher.

In this book, I first wanted to explain what artificial intelligence is. What do we mean by that? There are always those who say that it represents the use of a wrong word, or that it is a bad translation from English, or that it is impossible. But this is not the case, and with these introductory remarks, I will try to explain that it has a very precise meaning. Secondly, what existed before artificial intelligence? For a long time, people have been thinking about how to make machines that can simulate our reasoning. I will give some background on its history after presenting the definition of artificial intelligence. Finally, I would like to recount recent history, talk about generative AI, and conclude with the very complex question

that was asked of me, and which raises a great many questions: where is AI headed?

Without further ado, what does "intelligence" in artificial intelligence represent? Do you know what this intelligence is? The term is polysemic; there are at least four meanings to the word intelligence. The first is mind. So, is this what artificial intelligence is all about, building a machine with a mind? As a scientist, let me reassure you that the answer to that question is no.

The second meaning is ingenuity. Are the machines we build clever? Like a student at school, who is said to be clever. That means he is good at math; he has come up with original solutions and so on. Is this what intelligence is all about? We do not know. Machines are not always very clever, just as humans are not always very clever. It is a complicated question. There is IQ, but it is scientifically extremely problematic. A well-known French doctor, who prides himself on his knowledge of artificial intelligence, claims that machines will have a phenomenal IQ, higher than humans, but that is absurd. In fact, IQ is an arbitrary measure that assesses the sum of individuals' performance on numerous tests. It is standardized so that the average for the population is 100. But we do not really know what that equates to.

Thirdly, some pseudo-intellectuals claim that artificial intelligence is a mistranslation of English in expressions such as Intelligence Service or Central Intelligence Agency (CIA). AI is said to represent information flows, coding and so on. Rest assured; the term also exists in French. When you say that you are on good terms with someone, it means that you have a relationship with that person. Similarly, as a lawyer, you must know what "collusion with the enemy" means: it means leaking information... So, the meaning also exists in French, and the intelligence in artificial intelligence is not a bad translation from English.

The real meaning of intelligence in "artificial intelligence" refers to the ability to know, and more generally to all cognitive functions. What is its origin? It was the positivist philosophers in the late nineteenth century who tried to naturalize philosophical issues, particularly the study of the mind. To do this, they tried to use empirical methods of physical science, experimentation, to address the different mental faculties. In France, there was a philosopher, whom some of you may know from his book, "Voyage en Italie", Hippolyte Taine. In addition to his essays on aesthetics, he

has also written a large two-volume book, "De l'intelligence" in which he discusses the various mental faculties. I recommend it, even if it is a completely outdated work, because it is beautifully written and it talks about memory, reasoning, all our cognitive abilities. It is precisely this meaning that is used in artificial intelligence.

This leads me to one observation: intelligence is certainly not universal, contrary to what several people would have us believe. In popular discourse, it is often said that the intelligence of machines will increase with their computing power, so that suddenly they will surpass us, as if mental abilities could be accumulated. This is not the case. Intelligence is not a substance; it is a set of varied and immeasurable cognitive faculties; these faculties differ not only between species, but also between individuals.



Intelligence is not a substance; it is a set of varied and immeasurable cognitive faculties: these faculties differ not only between species, but also between individuals.

Functional brain imaging techniques are very helpful in demonstrating this, as they visualize brain activity. Here, in A, you have the image of a musician's brain in activity while listening to a piece of music. Second, at the bottom, in B, you have another brain with functional brain imaging of someone who is not a musician listening to the same piece of music. On each of these brains, the red and yellow areas correspond to the active areas that we can track in real time. You can clearly see that the active areas are not at all the same for the musician and the non-musician. In conclusion, if you listen to music, depending on whether you are very familiar with the music, whether you know it or not, your activity is not at all the same. So, the idea that there is such a thing as general artificial intelligence is based on nothing. I insist on this point: intelligence is not a substance.

This brings us to the second question: when did all this begin? A very long time ago. In the

17th century, Pascal invented a calculating machine. It was not a computer, just a calculator. Nevertheless, he said that this machine produced effects that were closer to thought than anything animals do. This intuition is incredible. However, he realized that there was no reason to believe that it had a will of its own, unlike animals. He wonderfully summed up in one sentence the question we are asking ourselves today, and to which we will return at the end of this presentation.

As a result, several attempts have been made to build machines that could simulate human reasoning processes. Another 17th-century philosopher, Leibniz, arithmetized logic, i.e., the laws of thought. He then built a machine to calculate this arithmetic of logic, still with this goal of stimulating reasoning. The project was never completed, but the intention was there. Later, in the early 19th century, Charles Babbage imagined a machine capable of executing algorithms, i.e., a computer. His assistant, Ada Lovelace, then envisioned that this machine could be used not only for mathematical calculations but also to process texts and music. You can see how this brings us clearly closer to artificial intelligence. Still in the 19th century, Stanley Jevons created a "logical piano"¹ to simulate reasoning. In the early 20th century, in 1912, a Spanish engineer named Torres y Quevedo built an electromechanical machine that played chess endgames, king versus king. You can see that these attempts were emerging gradually.

Closer to our time, in 1943, you can find two fundamental articles. The first, by Norbert Wiener, Arturo Rosenblueth, and Julian Bigelow, focused on teleological machines. The prefix "teleo" refers to the Greek word *teleos*, which means purpose, in this case the purpose of machines designed to achieve a goal. This article laid the groundwork for what we now call feedback and cybernetics. The second article, which is central and plays an extremely important role, was written by McCulloch and Pitts and dealt with formal neurons. What was the idea? It was to simulate brain activity with automata.

The first electronic computer was built in 1946. Before that, there were calculators. These early machines used telephone relay systems. We then tried to use these devices to simulate what we knew about our brain. As early as 1906,



The first electronic computer was built in 1946. Before that, there were calculators. These early machines used telephone relay systems. We then tried to use these devices to simulate what we knew about our brain.

it was already known that the brain was made up of several cells and that there were connections between these cells. These connections, known as "synaptic connections", are plastic, meaning they evolve over time, and this relates to memory and learning. So, there were attempts to simulate this using telephone relays. We will return to this later, as these models are becoming increasingly important today.

Machine translation began shortly after the advent of the first computers, in 1949, with early attempts by figures such as Warren Weaver, founder of information theory alongside Claude Shannon, among others. This was followed by the birth of the first project to build a translation machine. What was being done at the time? The approach resembled that of a poor student: taking one word at a time, trying to find its equivalent in the dictionary, then doing the same for the next word... Needless to say, it did not work. Since this approach failed, researchers turned their attention to context. Take the French word *manche*, for example. In the masculine form, you might say *manche de marteau* (hammer handle), right? But not necessarily. The problem is it could also mean *manche de gigot* (leg of lamb handle). So, a broader context is needed, which makes the task quite challenging. It gets even trickier with the feminine form *la manche*, which could mean a shirt sleeve (*manche de chemise*), begging (*faire la manche*), or a round in a card game (*la manche au bridge*), and so on.

Translation is a complex operation, with many obstacles to overcome, which I will not go into detail here. To illustrate the difficulties, let us look at pronoun ambiguities. Consider the sentence: "Take the fish and remove the scales, put them in the dishes." How do you determine the reference of the pronoun "them"? If you take the closest

1. "The logical piano", also known as the "Jevons Piano" in French.

plural noun, it would be the scales. Do you put the scales in the dish? We know that this is not the case, but how does the machine know? This set of difficulties led linguists in 1966 to say that it was necessary to stop and that it was impossible to do machine translation with this approach. Linguistic knowledge, a lexicon, grammar, and pragmatic knowledge (such as the fact that fish scales are thrown away) are essential to succeed in this mission. You will understand that they were completely wrong, but we learned this 40 years too late. And that is why the history of artificial intelligence is littered with mistakes and failures. When you ask me where artificial intelligence is headed, the truth is that we do not know, and we cannot know. Every time, we make up stories about what will happen in the future, and reality never matches what we imagined.

What happened next? Artificial intelligence was born in 1956 with John McCarthy, a relatively young man at the time, 28 years old, with a PhD in mathematics and a brilliant mind. Together with three other people, Marvin Minsky, Nathaniel Rochester—IBM's chief scientist—and above all the great Claude Shannon, the man behind information theory, he decided to elaborate a scientific discipline whose aim would be to simulate different cognitive faculties. Here is its definition.

It is a discipline based on the conjecture that all characteristics of learning and all other components of intelligence can, in principle, be broken down or described so precisely that a machine can be built to model them. In other words, intelligence is broken down into cognitive faculties that can be simulated. This has nothing to do with the idea that a machine can be animated by intelligence, or that it will be wise. It is simply a better understanding of all our cognitive faculties by reproducing them.

As Americans, they were pragmatic, and thought about utility, in particular the possibility of introducing this technique into technologies such as autonomous cars. And at the time, they were already brimming with ideas. The first robot models already existed in 1956. You can find archive footage showing Albert Ducrot presenting a small robot. It has sensors that send signals, which modulate its activity. This is what distinguishes it from a simple automaton that executes sequences of operations without taking the outside world into account. With artificial intelligence, they hoped to go further and simulate all mental faculties: reasoning — for example,

proving theorems or making logical deductions —, perception —, recognizing faces and speech —, communication, memory, learning, etc.

Artificial intelligence is an extremely broad discipline that has made considerable progress over the last seventy years. To give you an idea, here is a short summary of its history. As you will see, it did not all start two years ago.

Artificial intelligence was born in 1956. This birth was closely followed by the first achievements, particularly the first automatic theorem proving programs, which were undeniably successful. Right after that, machine learning—and I emphasize that machine learning was not created ten years ago—was born with the beginnings of AI, thanks in particular to a pioneer, Arthur Samuel, an engineer at IBM. He had built a machine that played checkers. He wanted to improve his program by trying to model the learning theories that came from behavioral psychology. Here, I am drawing Skinner's box, which you may already be familiar with. A mouse is placed in a box, and if it touches a certain lever or goes to a certain place, it is rewarded. On the other hand, if it touches another lever, it is punished, and we observe the time it takes to learn that it has a reward if it acts in a certain way and a punishment if it acts in another way. To simulate this learning process, we proceed by placing small virtual pebbles to mark out paths, and this method yields very interesting results. And we are still using these learning techniques. As you know, ChatGPT uses language models. However, to avoid deviations, it also uses reinforcement learning tools with human feedback. Instead of mice, we try to simulate moderators living in Kenya, but regardless, the principle is the same.

People were enthusiastic about the initial results. Especially Herbert Simon, a pioneer in artificial intelligence who would go on to win the Nobel Prize in Economics and was a specialist in social sciences, declared in 1958 that if machines were not excluded from chess competitions within ten years, they would become world champions. They should also be able to prove mathematical theorems and generate music of undeniable aesthetic value... Even back then, it was already being suggested that AI could be generative.

Nevertheless, after a few years, two problems arose. First, as mentioned earlier, machine translation: as we have seen, linguists claim that it cannot work, at least using the techniques currently

available. On another note, in 1965, a chess program was created that was beaten by a 10-year-old child. It then became clear that the ten-year deadline for seeing a world champion chess program would never be met. This led to what was called the "winter of artificial intelligence". This expression has been used many times since. For example, you mentioned 1992 earlier as the winter of artificial intelligence, but these regularly mentioned winters were not that cold. My first book, "L'âme machine" dates from 1990, and at the time, the subject was still very hot, I can assure you.

The first chatbots date back to 1965, the beginning of this so-called winter of artificial intelligence. Yet we still use them today. And shortly after, in the 1970s, there was "SHRDLU," the first natural language understanding program capable of interpreting simple sentences such as "take the green cube behind the red pyramid" and having a robot execute them. What is interesting is that they made a program that asks you to take the glass butt behind the pyramid. And then, a handful of robots will understand the instruction and take it. In France, in 1972, at the height of the so-called winter of artificial intelligence, a great mathematician, Alain Colmerauer, designed the PROLOG language, which allowed computer programs to be written using mathematical logic and executed with an automatic theorem prover. In short, during this period, work continued, and extraordinary results were achieved.

Then, in 1979, there was a rebirth of artificial intelligence. This was due to an observation made by pioneers in human intelligence. They realized that one of the reasons for these failures was that reasoning, and therefore intelligence, had been reduced to mathematical formulas. So, they pivoted attention on human intelligence.

Exchanges between psychologists, computer scientists and linguists gave rise to the cognitive sciences, which focus on knowledge. This led to studies on the organization of knowledge in human memory, to model it. This is known as knowledge representation.

This was followed by the creation of systems that model the knowledge of professionals in an attempt to be more efficient. These so-called "expert" systems have achieved extraordinary performance in the fields of chemistry, medicine, geology and many other areas of practical knowledge, which we could not imagine initially. In fact, at the beginning, the idea was that it would be easy to get machines to do mathematics, and there were doubts about the success of computers in disciplines that require intuition. In the end, excellent results were achieved in these areas.

Of course, this knowledge that we were trying to model with expert systems then had to be acquired, and it was more complicated. A new profession was born at this point, cognitive maieutics, which covers how to help experts explain the knowledge they use in their professional practice. The difficulty lies in the fact that experts, in general, do not know what they are doing. They do it well, but when asked how, they are unable to explain it. They omit certain steps and, sometimes, some lie when they mention what they should do, as learned in school, when, in reality, they do not proceed in this way at all.

This led some specialists, me included, to say that this knowledge should be acquired automatically through machine learning, based on data. Of course, this was the revival of machine learning in the mid-1980s. At that time, many machine learning techniques had been developed, which I will not describe here due to lack of time. One of these machine learning methods is formal neural networks, which were revived at that time, in 1986. One of the people behind the new learning algorithms for formal neural networks is Geoffrey Hinton, winner of the Nobel Prize in Physics in 2024.

This physicist revived the theories of statistical physics to study the dynamics of formal neural networks, which then enabled him to generalize the learning procedures devised in the 1950s and achieve the success we saw in the late 1980s. The problem was that, at the time, machines were very slow and therefore unable to learn efficiently. Neural networks consisted of 10 to 20 neurons, and the results were not very convincing. Other



Exchanges between psychologists, computer scientists and linguists gave rise to the cognitive sciences, which focus on knowledge. This led to studies on the organization of knowledge in human memory, to model it. This is known as knowledge representation.

techniques were gaining momentum, such as the construction of decision trees or random forests, Bayesian networks, and SVMs (Support Vector Machines), so that from 1992 onwards, the use of formal neural networks was discontinued, except by Yann Le Cun. He was working in the United States and took the opportunity to continue. As machines became very powerful, instead of working with small neural networks, he was able to work with very large neural networks comprising 15 layers of neurons. These were called deep neural networks because they have many layers. The results were breathtaking, but I do not have time to go into detail.

You know how this story goes. With these deep neural networks, we have produced generative AI. Let me briefly remind you what this consists of. The idea is that these deep neural networks will enable you to capture the spirit of the language. It is a rather mysterious idea. It is the invisible relationships between words that we can learn automatically.

Let me come back to what I was saying earlier. In their ALPAC report, linguists said in 1966 that if you want to process languages, you need linguistic knowledge. You seemed convinced. But in fact, among engineers in the 1990s, who were themselves using machines for speech recognition, some were much less convinced than you were. One of them even claimed that when he fired a linguist, the performance of his system improved... This was very derogatory for linguists. And yet, it turned out to be relevant.

So, what is the idea? It is to extract this linguistic knowledge, but not with words, only with numbers. You input sentences into huge artificial neural networks and try to recover the same sentences at the output of those networks by adjusting the connection weights of the artificial neurons. Behind this approach lies the old concept of distributional semantics, which holds that meaning is tied to the context in which words are used. This is the notion of distributional semantics. Meaning is linked to the context in which words are found. That is what these systems are trying to capture, and it is remarkable.

In fact, we are going to build machines that can perform a variety of tasks, ranging from natural language processing to text summarization, or even the development of question-and-answer systems. However, the most extraordinary thing is that when you give them a sequence of words and ask what the next word will be, they find the

most likely word to follow, then the next, and so on. You are going to say this is completely silly, but from the results, it seems to make sense. That does not mean it is true, but at least we understand what he is saying—it makes sense. He does not understand, but that is another matter.

And we are training this based on hundreds of thousands of books, which is dizzying. It is on the order of a tenth of the "Bibliothèque nationale de France", which has 14 million books and printed materials in its catalog. These massive neural networks have on the order of 1,000 billion connections, which is equally staggering. Through these trainings, they learn the relationships between words. The texts that come out—you have all tested them—look well-assembled. There is no reason for it to be true, even if it is the case from time to time. Most of the time, it is commonplace. Sometimes, however, it can be completely false.

I have a little story here to share, which I find delightful. One of my readers asks ChatGPT the following question: "‘Tu y pourras reconnaître l'idée De la beauté qu'en ce monde j'adore.’ By which author and in which sonnet can we find this abstract?". ChatGPT replies that it is from "Le Lac" by Lamartine. This is rather funny, as Lamartine never wrote a sonnet, and "Le lac" clearly is not one. And he gives the text, which we compare with the real one. The first stanzas are correct, then he added two stanzas to include the famous verse. He has written and invented fake Lamartine.

The same reader asks Gemini the same question, which answers correctly that it is from Du Bellay's "L'olive". He then asks if he can give him the text of the poem, to which Gemini replies that he cannot, as it is too long. But a sonnet is very small. What is more, and this concerns you lawyers, he adds that it presents legal problems. The reader then informs Gemini that ChatGPT had told him it was by Lamartine. He then says that this is a very interesting literary enigma, and that sometimes authors take inspiration from others, plagiarizing them and that, perhaps, Du Bellay had plagiarized Lamartine. The reader retorts that Du Bellay was born long before Lamartine. "You are right," says Gemini, very compromising. Of course, everyone knows, and he more than any other, that Du Bellay is a 16th century author and Lamartine a 19th century one. It is impossible for Du Bellay to have been inspired by Lamartine. The illustration is humorous, and you must laugh

about it, but the main thing to remember is that it is not to be trusted.

Finally, where is AI headed? There are many fears, and I will not go into them all here. Some are baseless. General artificial intelligence, as I said, does not exist. Strong AI—the idea that there is consciousness behind the machine, the so-called existential risk in Elon Musk's words, the notion of machines taking over—makes no sense and is based on nothing at all. There are dangers, of course, especially the replacement of humans by robots. But what does this encompass? These are questions that need to be asked. There are, however, questions that are closer to today's concerns, and which, as a jurist, we need to work on: the question of human responsibility, bias and discrimination, these are real issues.

How can we proceed? We can declare a moratorium, decide to stop everything, but there is no chance of that working. You could have a red button, but that is complicated. In the aeronautics industry, for example, we have realized with hindsight that in conflicts of authority between men and machine, very often it is men who make mistakes, and it is better for the machine to take the decision. Do we need regulation? As lawyers, I am sure you would say yes. But would regulation be enough? That is one of the questions we will come back to later, even if I am rather reserved on the subject, and I believe there are others here who share my concerns.

Finally, I would like to end with the "bewitched charity", an expression I introduced in my book, "Le mythe de la singularité", taken from Rimbaud's poem collection "Une saison en enfer", in which he mentions the bewitched charity of the infernal spouse. I would like to give some illustrations of the bewitched charity from the big Internet players. In 2020, IBM, which was very well disposed, announced that it was withdrawing from the facial recognition market because the technology would be used for racial profiling—which is commonplace today—and called for police reform. At the same time, a report on facial recognition was drawn up for the CNPEN, citing IBM. Shortly afterwards, a colleague shows me articles from Liberty Investigation and The Verge which reveal that, in September 2023, IBM signed a \$69.8 million contract with the British government to create a biometric platform with facial recognition functionalities. This gives you an idea of what these major Internet players are up to, as they change sides on a regular basis.

Another example: one of the endorsers of the call for a moratorium is Elon Musk. Just fifteen days after, he finally says that ChatGPT is very dangerous, because it is "woke", and that he will tell the truth. He then decided to launch a company to create the TruthGPT system, which he described as a "beacon of truth in a dark world". For a second example, we must go back to January 7, 2021, when Mark Zuckerberg had decided to ban Donald Trump from his social networks and close his accounts. Four years later, on the anniversary date, January 7, 2025, he said he had been wrong, that the most important thing is in fact freedom of expression and that, not knowing the truth, we should not risk being unfair. There is a plethora of examples, but I am not going to give them all.

To conclude, I would like to quote a text I particularly appreciate from Étienne de La Boétie: "Theater, games, farces, spectacles, gladiators, curious beasts, medals, paintings and other drugs of this kind were for ancient peoples the bait of servitude, the price of their ravished freedom, the tools of tyranny. This means, this practice, these enticements were those employed by ancient tyrants to lull their subjects into submission. [...] But today's tyrants are no better: before committing their most serious crimes, they always precede them with a few pretty speeches about the public good and the relief of the unfortunate."

Decoding AI means seeing what is inside the machines, inside the code, and identifying what dangers might be lurking there. There is a science-fiction writer, Ray Bradbury, who was asked if he feared robots. He replied that no, it was not the robots that scared me, he said, it was the men behind the robots.

I believe he was right. I think that digital giants are what we need to be wary of today. Google's "Don't Be Evil", Elon Musk's takeover of Twitter to preserve freedom of expression, the end of moderation by Meta, always concerned about freedom of expression, Elon Musk's moratorium on artificial intelligence... These are just a few examples. I think these are the real issues we should be talking about today. ■

Fireside chat:

Speaking the language of generative AI

- What are the latest advances in generative artificial intelligence?
- Understanding the conceptual foundations of artificial intelligence models to optimize their use and guarantee their governance
- Automation must be achieved by giving it meaning

Yannick MENECEUR

Magistrate and associate lecturer at the University of Strasbourg's Faculty of Law, Political Science and Management

Murielle POPA-FABRE

Expert at the Council of Europe on issues of artificial intelligence regulation and data protection

Thomas LAMPERT

Professor of Artificial Intelligence and Data Science at the University of Strasbourg

Yannick MENECEUR: We have seen since late 2022 public space is saturated with the achievements of ChatGPT and generative AI. And while Microsoft researchers believed that they had found sparks of general intelligence in the most recent models of the time in 2023, others continue to search for them today. This is why it is essential, both for professionals and citizens, to take an objective look at these latest technological developments. I would like to once again thank Jean-Gabriel GANASCIA for his brilliant presentation, which helps us see things more clearly. Murielle POPA-FABRE, you are an expert to the Council of Europe and to the United Nations on matters of artificial intelligence, and a doctor of language neuroimaging. You have written for several national newspapers, taught at Cornell, and regularly speak at conferences like today's to skillfully popularize the concepts driving this generative AI.

Thomas LAMPERT, you have also served as an expert to the Council of Europe. I am reminded of the American campaign we experienced

together, though I will not go into detail, or only under constraint. You are a professor of Artificial Intelligence in Science and Data at "Télécom Physique" here in Strasbourg, and you also tenure the Industrial Chair in Data Science and Artificial Intelligence. You are specialized in image and time series analysis for various fields of application, notably in medical imaging and remote sensing. Thank you very much for being with us today.

We will take 30 minutes to discuss the latest developments in generative AI. Some groundwork has already been laid, thanks to Jean-Gabriel GANASCIA, but we will now delve deeper into some of the concepts, which will serve as a foundation for later discussions. It is indeed important to establish a solid base of understanding to avoid any misconceptions or fantasies.

A very first question, following what we have just heard, and I am sure you have something to say about it: is a "thingummy" like ChatGPT intelligent or not?

Murielle POPA-FABRE: I come from cognitive science, so I will not venture into defining intelligence. Instead, I will show you how the question you ask your conversational agent, such as ChatGPT, is visualized by the machine.

To achieve this, we need to study the vectors, which are the sequences of numbers that translate words for the machine that Jean-Gabriel mentioned earlier. They are represented spatially, as these digital objects are what generate your response and play a role in the probability calculations that produce it. First, what you get is kind of an illusion of answers—it is just the statistical continuation of the words in your question. What is captured in these probability calculations



Yannick MENECEUR (moderation), Thomas LAMPERT, and Murielle POPA-FABRE.

is the journey of words. Thus, the real question of distributional semantics, and of the simulation of language generation as performed by conversational agents, is to ask: with which other words do the words you used in your question most frequently travel?

For instance, if you take an economic corpus, the words “economy”, “rate” and “growth” will be highly likely to appear together. This is the result of processing tens of thousands of web pages on the topic of economics, which show that the term “growth” is often linked to “rate” or “economy”. So, for a mechanical arm programmed to construct probable sentences, the fact that these words often appear close to one another makes it easy and efficient to respond to a user’s question about the economy. Behind all this lies a major issue: the representation and simplification of the world. For example, if you go to the BNF (National Library of France) and gather all the articles written about degrowth to feed into your GPT-2 training set, while removing all those related to growth, then the machine will most likely talk about “degrowth” when you ask it a question about the economy.

If this conference had taken place four years ago, when the field of interpretability was not yet very advanced with this type of neural network, I would have told you that a form of intelligence lies in the geometry we can perceive from the

proximity between these words. The similarity between these words would then stem from a calculation. Indeed, the machine determines the proximity of word sequences to calculate their probability of being associated in a sentence. Each word is hence characterized by a kind of checkboard pattern.

For example, “cat” and “dog” often appear together in the same sentences, so it is more likely to determine similarity within this grid than with the one representing the word “chair”. But if there were an animal only found on chairs, its grid would closely resemble that of “chair”, because in the data mixture used to train the model, it would almost always appear only alongside the word chair. This is what we can call a statistical representation of language, based on very large volumes of data, to represent one of the functions of language that borders on semantics.

Since the middle of last year, very interesting results have been achieved thanks to the development of a field called Mechanistic Interpretability. One study in particular managed to capture what happens inside the neural network at the heart of ChatGPT. By analyzing the training process of a large language model with 70 billion parameters midway through its learning, researchers from a ChatGPT competitor were able to extract what had been condensed from training on millions of textual data. Using a

technique called dictionary learning, they created a map of what is represented inside the neural network. To use a familiar metaphor: it is as if we had a giant vacuum cleaner that condenses the relational information between words.

For instance, researchers have documented what can be visualized regarding the term “inner conflict”. If you dive into this visualization, you find everything that is considered similar and closely related, always within the same dynamic. However, this is no longer just the representation of words; you are inside the neural network, within what the training phase has allowed it to condense from millions of sentences in its training corpus. The expression approaches terms like “impossible choice”, “curious paradox”, “impossible desire”, “dilemma”, the question of guilt, or even the situation described in the book “Catch-22”. That novel deeply impacted the United States with its depiction of a soldier trying to leave the front during the Vietnam War. Over time, through constant condensation and learning from many texts, what is aggregated inside the neural network reaches a level of sophistication that we can now interpret thanks to the field known as Mechanistic Interpretability.

If you want to test this on a smaller model, you can go directly to Neuropedia, which condenses what is inside the Gemma model, Gemini’s open-source version. All of this to say that we are reaching a point where we can quite literally have a “right to look” inside — an observability of the sophistication condensed within neural networks. This can collectively serve us in the future to govern these machines and their outputs.

Yannick MENECEUR: Thank you very much, Murielle. I will attempt a summary. My understanding is that we have managed to make language calculable and mathematizable, and today we have a set of mathematical, probabilistic, and statistical methods to capture the texture of language.

Thomas, we will not leave you out of this discussion, with your perspective. I have the feeling that, in the end, the revolution is just around the corner and that we are finally going to see this much-talked-about general intelligence, even though Jean-Gabriel described it as quite an abstract concept. For example, we have seen Elon Musk’s robots pour beers and do incredible things. The question is, is there really an ongoing revolution with ChatGPT, if it indeed is one?

Thomas LAMPERT: The set of neurons we have been discussing this morning, is after all the same type of neuron that was discovered in the 1950s. There is nothing revolutionary about this part; it is more the organization that is revolutionary. Truthfully, what has changed recently, on the one hand, is the computational powers of Graphics Processing Unit—GPUs—, which we use for gaming, but which we can harness advantageously in the context of artificial intelligence. On the other hand, it is also the amount of data available and the way data is used that constitutes the revolution.

Before the 2010s, we had to label all the data. We had to say, “this is an image of a cat”, “this is an image of a dog”. It took a lot of time and human effort and limited the potential for learning. With this type of model, we can use any unlabeled data. If we imagine that we hide a word and ask the model to recreate it, this is a task already present in the data. We do not need to label anything because we remove the word. We know the word, the result. Thus, we can create models and train them in this way.

This means that between all these lines, there are a lot of numbers, multiplications, because we need to multiply the input of this representation that was explained. We use large amounts of data to guide these numbers toward the right result. After that, we get hold of the model that is ready to be used. The revolution is this. It is not the models themselves. Of course, there is an evolution of the model, which will become more complex over time, but that is the result of these advances. There are new mechanisms, but not a major change in the type of model.

Yannick MENECEUR: We have been able to make things work that were there before, but more efficiently, in other words. Now that we have gone back to the basics, let us take a reroute to discuss current events.

Personally, I must admit that I am having a hard time keeping up with the pace of these developments today. Last year, I bought another “thingummy”, the Rabbit R1. This small red-orange box was supposed to be able to perform actions after giving instructions out loud. The promise was that we had a tool that would allow us to automatically perform everyday tasks. I must admit, it is currently serving as a paperweight. This brings us to the topic of large action models, or LAMs, about which I have read a lot,

but they may not necessarily be a very precise representation. Help us clear things up on this LAM concept. What exactly are they capable of?

Thomas LAMPERT: It is a model that can perform actions. To simplify, it can control the graphical interface of a computer to accomplish a task. This is the first type of model that can do this. In fact, there are two things to say. First, within LAM, there is an LLM (Large Language Model), which is an identical model trained with language. Secondly, we slightly modify the model based on the task for which we want to use it. This part involves controlling a computer. In the theory of intelligence, we can say that planning is a very important task that demonstrates our intelligence, and we are starting to see models perform this task.

But once again, it is not a revolutionary model; it is a first model to which we add another box inside to adapt it to our task. For some, to achieve artificial intelligence inside a computer, it would require interaction with the world around them.

Murielle POPA-FABRE: There are numerous theories of intelligence, and it is true that the interface with the outside world is important. There are also ways of conceiving intelligence as embodied, and part of the field of robotics seeks to draw inspiration from these cognitive theories. We seem to be entering a new cycle in the development and deployment of artificial intelligence where, instead of relying on a single, monolithic neural network, we are beginning to create composite systems. This is something we have already seen in practice with the arrival of RAGs². Retrieval Augmented Generation has introduced the possibility of conducting research within a defined set of documents to feed the language model, using it as a writing engine, while ensuring that it alters as little as possible the input data.

What I find interesting about this new wave of automation is that economists are now speaking in terms of cognitive automation. Indeed, what we are increasingly trying to reproduce are steps—the ability to predict in a model how the human brain works, which is quite modular. This is what we have called the modularity of the mind. The idea of building increasingly modular systems thus carries with it stronger automation, covering a broader range of human activity, along

with a certain form of autonomy. For instance, in some action models, the decision-making process involves not deciding and instead letting the model itself choose between options. Despite building a more explicit architecture, less probabilistic than the simple neural network, we now find ourselves in a situation where we must ask how we observe what is happening—and how we control it.

Yannick MENECEUR: I wanted to discuss with you this concept of Large Action Model because these are terms that have been coming up recently. I have also heard about other concepts that seem important in terms of current developments, such as agents and design models... What interests me, Murielle, is whether all these elements—LAMs, agents, design models—will help improve the performance of LLMs, which unfortunately tend to invent things, as we saw earlier, especially in specialized contexts like law. This is also what interests us today: how to make this type of technology useful, efficient, and operational today in the field of law.

Murielle POPA-FABRE: I would always respond by saying that in the field of law, we have many professions that also have quite different ways of proceeding. Therefore, it is difficult to speak out on all the professions within law. I would like to thank the IERDJ, which I had the opportunity to work with on the question of judicial decision-making and generative AI.

Yannick MENECEUR: This work is available on the IERDJ website, by the way, if any of you wish to consult it.

Murielle POPA-FABRE: It is through this approach that I started reflecting on the process followed by this eminent profession of law. I will now talk to you about a Large Concept Model, developed recently by a Meta lab. I found the reasoning behind it very interesting, based on law and the fact that the judge represents an interface between the real world and the rule. Still in this idea of finding the decoding key, in this LAM, the minimal unit is not the word or the token, as I showed you earlier, it is a concept, meaning a small part of a sentence.

What I find interesting is this possibility of no longer relying on elements that have no interpretable meaning for humans but rather starting to

². Retrieval Augmented Generation.

train on elements that are understandable. In this way, we are building something that approaches a kind of logic, always based on a linguistic element. We will surely see more and more attempts to model increasingly sophisticated, multimodal things, like pragmatics—hence the interest in keeping Rabbit R1 when they will change the algorithm. The more we follow these directions and go higher, the more we will capture interactions and elements that, starting from meaningful building blocks, will allow us to construct scaffolding with bricks that make sense. These technologies and techniques will be able to automate quite basic things, in my opinion, and avoid that somewhat loud aspect that you described, Yannick.



The challenge ahead, is the question of observability. [...] We are going to see systems that automate much more, and we will need to find ways to observe what they are doing, both to govern them and simply to improve them. [...] The key question is: where is the human essential and prioritized in automation?

Yannick MENECEUR: This parenthesis is soon coming to an end. We are now approaching the final question, which I would like to keep very open and forward-looking. Your challenge will be to answer it briefly. Ultimately, are these LLMs, along with the full range of technologies we are using, reaching a ceiling, a limit? What, in your view, could enable us to move forward in a truly decisive way?

I am thinking of Yann LeCun, when he fought alongside his colleagues Yoshua Bengio and Geoffrey Hinton to promote deep neural networks, at a time when everyone was swearing only by support vector machines or other approaches. Are we not now reaching a ceiling with these current technologies? What do you think would truly be a game changer to go beyond?

Thomas LAMPERT: A ceiling? I do not know. I believe everyone can add new building blocks, as we have explained, and new possibilities

emerge with the model. Personally, from a technical point of view, it has already been said that we need to combine these logical parts with the connectionist ones—in other words, hybridization—to demonstrate this, and we can see that this is starting to emerge from this characteristic. Thus, there is something to be done from that perspective. I am not sure whether it will happen through increasing the complexity of the models, or if we will need to shift toward a different combination of things to reach that stage.

Murielle POPA-FABRE: The real bottleneck, the challenge ahead, is the question of observability. The tech world, and venture capital, which is heavily investing in it, are now developing a new layer in the well-known “AI stack”. We are going to see systems that automate much more, and we will need to find ways to observe what they are doing, both to govern them and simply to improve them. So, there is this dual role: keeping humans as observers and guardians of the machine, but also as contributors to its refinement. In fact, these observability layers often serve this dual purpose: monitoring what is happening within the system and enhancing it.

This final image can be found within agentic systems. The human is positioned somewhere in this orchestration between language models and research supervisors, in the chatbot interface. The key question is: where is the human essential and prioritized in automation? Since this type of automation is cognitive, it also requires meaning, and it is therefore impossible to automate broad areas without meaning. I am convinced that the major shift ahead will be in defining the interface between humans and these systems. The real journey ahead lies in determining the role of the human.

Yannick MENECEUR: Since the graphical interfaces of the 1980s, we have not really seen a true revolution in how we interact with machines. Now, it seems, something significant is happening, and this will undoubtedly continue to fuel our conversations—and perhaps, without getting ahead of ourselves, continue to inspire our discussions together, maybe even into next year, which I truly hope. ■

Round table:

Will AI transform freedom into destiny?

- Analyze the societal and political implications of artificial intelligence on our individual freedom
- Assess the dangers of algorithmic determinism, which could reduce the field of possibilities
- Promote collective, democratic reflection on the integration of artificial intelligence into our societies

Moderation: Harold ÉPINEUSE

Deputy director of the IERDJ

Maxime DERIAN

Doctor of digital socio-anthropology

Antoinette ROUVROY

FNRS researcher at the "Centre de Recherche en Information Droit et Société" (CRIDS) at the University of Namur

Bilel BENBOUZID

Senior lecturer at the University Gustave Eiffel, "Laboratoire Interdisciplinaire Sciences Innovations Sociétés" (LSIS)

Sabine VAN HAECKE LEPIC

Lawyer and Doctor of law

Harold ÉPINEUSE: A round table entitled "Will AI Turn Freedom into Destiny?". This is a trick question, and we have not asked the speakers to answer it directly. It is a quote from Antoine Garapon, taken from his book "Justice digitale" (*Digital Justice*), working at the time on the issue of so-called predictive justice. We found it particularly relevant to illustrate the themes we will be exploring with our panelists—namely, the societal and political impact of AI on our lives. We will benefit from the insights of our four guests, who come from diverse backgrounds, in keeping with the promise of a multidisciplinary perspective.

Joining us is Maxime DERIAN, a socio-anthropologist from Luxembourg with a background in law. You are a researcher and consultant, and

lead the techno-realism movement, which offers updates and reflections on digital issues. Also joining us from abroad is Antoinette ROUVROY, from Brussels. You are a legal scholar, and your work in legal theory is well known, particularly on general normativity and algorithmic governmentality. You are a FNRS researcher, the Belgian equivalent of France's CNRS, and you work at the University of Namur. Also with us is Bilel BENBOUZID, who comes from Paris, whose background perhaps best illustrates multidisciplinary. You began with an urbanist, technical, and engineering profile, and are now a sociologist specializing in the sociology of science. You are a lecturer at Gustave Eiffel University and will share your broader perspective on these developments in AI. Finally, another legal expert: Sabine VAN HAECKE LEPIC, a lawyer, arbitrator, and mediator, Doctor of law, particularly in the philosophy of law. You will bring yet another distinct lens to this topic of AI. To begin with: when was your first encounter with AI?

Bilel BENBOUZID: My first encounter with AI happened about ten years ago, with the use of machine learning algorithms by police forces in the United States. Police patrols were using models to predict the time and location of future crimes. Gradually, this led me to explore the political implications of AI.

Maxime DERIAN: As for me, I became interested in the history of cybernetics to better understand all these new machines that were emerging. As a fun anecdote, I had a research residency in Japan around 2008, in a Precision and Intelligence department that was working with machine learning. There was a server there that trained neural networks. I could hear the engine whirring while I worked, until one day it crashed out from exhaustion; or rather it broke down. That gave me a very tangible, material perception of the object during training. So, during my first use of ChatGPT, I had this reflex to avoid asking too many questions, just to save electricity and avoid overheating the server, due to that materialistic awareness that I remembered. But of course, the level of complexity and power of these machines has since evolved significantly.

Sabine VAN HAECKE LEPIC: For me, it was while writing an article with Hervé Lécuyer about autonomous cars in 2018. The configuration of the decision-making field required asking ethical questions, such as “If the car has a technical malfunction, should it run over a baby or an elderly woman?”. These questions deeply moved me.

Antoinette ROUVROY: I have never truly “encountered” artificial intelligence—that is, systems that are both “genuinely” intelligent and “genuinely” artificial. What I have encountered, however, since I began working a little over 25 years ago on the relationships between juridico-political normativity, the normativity of living systems, and the normativity of statistic and algorithm, is a series of “turning points” (the genetic turn, the ontological turn, the computational turn, etc.). Each of these shifts repositions the privileged coordinates through which we perceive and anticipate geo-bio-physical, physiological, social, and economic phenomena, etc. The computational or algorithmic turn, by making digital data the privileged, if not exclusive, coordinates for modeling and modulating “reality” and now even for imagining it (through the artificial imagination of generative “AI”), is merely the latest in a series of such displacements. Each of these turning points also involves a set of reductionisms and gives rise to significant blind spots.

The Big data phenomenon, the increase in computational power, the abandonment of the deterministic metaphysics that still underpinned the work of the founding fathers of “social

physics” or statistics, such as Laplace or Quételet, and subsequently the rise of data sciences and computer engineering, have gradually transformed the “governmental effects” of these statistical practices. This transformation led me to orient my research toward algorithmic governmentality—that is, the idea of a (dis-)ordering of the world that relies less and less on subjectively and collectively identifiable and contestable norms, and increasingly on the agile detection of “patterns” within vast volumes of data.

Harold ÉPINEUSE: Thank you very much, this helps us better understand your point of view on the subject we are discussing this morning. I am going back and linking this with the documentary we watched on the work of Michał Kosiński in psychology. It would be interesting to know how you react to his statements and to share your different epistemologies, to speak a little pompously. What are, for you, the key elements to keep in mind when considering the societal and political impact of AI? We will give you the floor successively, starting with you, Maxime.



iHuman: L'intelligence artificielle et nous,
Tonje Hessen Schei, 2019

<http://bit.ly/42nbDHw>



Maxime DERIAN: Michał Kosiński's video inevitably caused a reaction within me, as well as his publications, as I imagine it did for you. It is thought-provoking because it immediately reminds me of phrenology, the math bump, the “thief’s head”. There is the question, not necessarily of the one of racial profiling we presume, but also of categorization. And fundamentally, when we categorize a human group, people generally tend to conform to the label imposed on them. Of course, there can be reversals of stigma,



Harold ÉPINEUSE (moderation), Bilel BENBOUZID, Maxime DERIAN, Sabine VAN HAECKE LEPIC, and Antoinette ROUVROY.

and a multitude of events can occur. But generally, if we take a group, for example, and start saying “you are part of the blue group”, “you are part of the green group”, a blue identity will slowly form, as well as a green identity.

And so, starting to divide human groups by saying “homosexual”, “not homosexual”, “right-wing”, “left-wing”, “I like dogs”, “I like cats”, “I do not like animals” ... We can create any type of division, it does not matter, it is very dangerous. Maybe it works, and it is proven that there is really a sexual behavior corresponding to a face, as this researcher claims, or maybe not. But just this act of categorization is playing with fire, and so, from an ethical level, I immediately raise the red flag.

In general, personally, my techno-realistic approach is to listen to everyone, examine all points of view, and analyze situations on a case-by-case basis. Profiling technology can be problematic in one context and not in another. For example, between the prison environment or the street. These are two different worlds because there are not the same legal environments for the individual moving within them.

But still here, I directly raise the red flag because it is dangerously ambivalent. Kosiński highlights this issue and proclaims himself a

whistleblower. And this is what seems truly disturbing to me because, in his figure of authority, he should oppose it and assert that this use is inappropriate. When we examine the uses of the AI Act in Europe, the question of allowed or prohibited applications arises, including social scoring—evaluating people based on their quality as citizens and their degree of obedience. Such a practice seems unacceptable. In Europe, this system is anyway prohibited because it does not comply with the provisions of the AI Act. However, his role should be to oppose these practices first and foremost, which he does not do.

This reminds me of Jaron Lanier, a computer scientist who, for at least the past fifteen years, has somewhat tempered his technophilic enthusiasm. One of the founders of virtual reality and the concept of the avatar, he now expresses a certain disappointment with some directions in research. He particularly regrets the funding of certain studies, such as those focused on hacking or cracking the security of a pacemaker, for example. This could lead, after several years of work, to a publication explaining how the security was successfully breached, opening the door for cybercriminals who might use it for malicious purposes. Of course, science and innovation should not be excessively limited, as

that would go against the spirit of our discussion today, which is to promote innovation and find new avenues. However, there still needs to be some guidance, a framework, and this is what I find disturbing once again in the philosopher's remarks, though he has certainly gained notoriety in the process.

To turn this page, I must say that my discussions with Yannick MENECEUR about the ethics of AI have had an influence on me. Until now, I tended to approach AI through the lens of the biometric system, phrenology, an approach where things are categorized as "good" or "bad". My evaluation relied on reactive, assessment-based logic, where the goal was to judge whether something was good, more or less, depending on the context. But with some perspective, this approach seems almost trivial in the face of the phenomenon. If we consider technological development as a flood—like Michel Tibon-Cornillot's concept—a nearly unstoppable wave, then constantly asking "is it good? is it not good?" is like trying to bail a tsunami with a teaspoon.

And so, after discussing with Yannick, I was convinced that a teleological approach to AI ethics could be a promising path. In other words, within the Council of Europe—if I understood correctly—there is this idea that AI should align with the cardinal values of the fundamental principles of democracy, the rule of law, and human rights. My first reflex was to think that these notions were too vague, too general. I wondered: how can we navigate with such vast concepts as human rights or democracy? And yet, this approach intrigued me. Indeed, I realized that returning to the previous state, evaluating each technology by asking whether it was good or not, was no longer suitable for what is occurring now.

That is when Hans Kelsen and the hierarchy of norms came back to me. We have general norms and principles of law, from which more and more specific norms flow, which can even define things like the size of door frames in new buildings. So, I go from a very specific case to something much more general, with a structure being created. It is a sort of grand legal cathedral that might need to be built. Following this logic, by breaking down these three major principles—human rights, democracy, the rule of law—I identified seven important issues. So, this is a step forward, although it is still an ongoing reflection, a work in progress.


Firstly, and this may seem a paradox, it is the necessity not to unnecessarily hinder innovation through regulation. And be careful, I am not saying that regulation is not needed. In Europe, our regulatory approach is often criticized, sometimes excessively, even to the point of self-flagellation. However, regulating is not about stifling for the sake of it: it is about setting safeguards, framing, and preventing excesses. It is precisely this approach that allows us to avoid doing anything and everything. Especially since we do not know, and Jean-Gabriel Ganascia may confirm this, where AI is going. Since no one knows its future, having a few safeguards seems necessary. All this to say that the first element will be to overcome technological obstacles, meaning creating appropriate regulation to stimulate innovation without constraining it too much, but with a minimum of safeguards. However, we must not prevent innovation, because otherwise, we will find ourselves dependent on Chinese, American, or foreign technologies.



In Europe, our regulatory approach is often criticized, sometimes excessively, even to the point of self-flagellation. However, regulating is not about stifling for the sake of it: it is about setting safeguards, framing, and preventing excesses.

The second issue is developing ethical technologies. It is a shortcut to use the word "ethical" here, as this term is complex and polysemous. It is not just a code of conduct that could be integrated into software, or written into a code, it is a process of reflection in collegiality, on a case-by-case basis. Ultimately, it is somewhat the role of the law, as the judge, to consider the context, to qualify facts legally, and to confront viewpoints. We still need to align what is being done at the technological level with what we desire at the social level. That's why the creation of connections, even hybridizations, between law and computer science seems beneficial to me.

I therefore used the term “algo-constitutional”, since in law school, the first courses taught are constitutional law within public law framework, and then, there is also private law. Perhaps we should create a new course, hybridizing both law and computer science, since Professor Lessig—a former techno-realist from the first school of thoughts—has told us that “Code is Law”. The algorithm is embedded in our reality, as demonstrated by the example of Facebook and its moderation. As soon as the policy of an algorithm changes, this political change is reflected in the social practices of billions of users, which is still extremely concerning.



As soon as the policy of an algorithm changes, this political change is reflected in the social practices of billions of users, which is still extremely concerning.

The third issue that arises is communication. We need to ensure a form of transparency, explainability, and a certain level of education about AI. A large part of the population is not sensitive to it and simply does not care or sees AI through a fantasy lens: either a Terminator designed to kill us all, or a new paradise, a new Eden. However, if we are called to vote on this subject one day, the voters must, at the very least, be informed.

A fourth, fundamental point is the environment. Earlier, we discussed the limitations of LLMs and the scale effect. More data requires more energy, whether nuclear or coal, to always achieve more. There is a constant need for more in this hyperbole toward an unattainable infinity. This seems unreasonable. So, it is not a question of excess, but of measure and environmental limits. The hedonistic approach of some, currently, of overconsumption, which relies on a prospective idea of nuclear fusion and resources from asteroids, represents “plans for a comet”, as we would say in French, a fantasy. These plans should not numb our vigilance; we must think of

the limits of environmental resources as responsible citizens.

Fifth, cybersecurity management is alarming. I place corporate cybersecurity and mental cybersecurity on the same level, that is, attacks on people's minds through disinformation, fake news for example. The manipulation of children and young people under 25 particularly concerns me, especially since they are growing up with increased proximity to chatbots. A certain psychological confusion could arise, with problematic anthropological repercussions, ultimately placing humans and machines on the same level: reifying people and humanizing machines.

Moreover, for institutions that use an LLM to function, the developers have a panoptic view of what is happening in real time within that institution, or even across all related institutions. This grants colossal power to the provider of such a tool, this cognitive assistance. And so, on this point, we must be very vigilant.

Finally, the last two areas are distinct, yet closely related. The first is inclusivity: going beyond the perspective of a dominant elite or a single school of thought. Techno-realism is about listening to everyone, including those who are skeptical or resistant to AI. It is just as important to hear their fears and concerns as it is to listen to AI advocates, and to help free both from their respective biases. It is not just about hearing them out but about thinking carefully and making decisions accordingly.

Inclusivity also means considering, for instance, gender issues. In Japan, for example, I met a researcher while we were working on designing exoskeletons. We created one for a large-framed man, one for a small-framed man, but none for women—even though the exoskeletons were meant to support an entire sector of the workforce. These are the kinds of biases—including those related to disability—that we must be careful not to overlook.

Finally, a last concept, difficult to define but closely related to inclusivity without being identical, would be that of non-imperialism. This may sound like a pious wish, but from an anthropological perspective, we must acknowledge the emergence of a hive-like society, driven by our use of cognitive prostheses. To summarize: for millennia, humanity was organized into communities and tribes. Then, with the Enlightenment (the “Lumières”), we shifted to a model based on

the social contract, where the individual became the central reference. While inequalities between individuals remained, the individual took precedence in this model.

Today, we may be witnessing a new shift: a return to unprecedented forms of community, swarms, structured through cognitive prostheses. These technologies act as intermediaries, “men in the middle”, in our social relations. They are neither neutral nor inconsequential. In truth, these tools are secreted—to borrow an anthropological term—by specific societies, like the United States or China.

This raises an important issue, because there exists, potentially or it is already the case, a form of imperialism. To illustrate this in very concrete terms, I am thinking of tokens. There is a real economic dimension here: tokens function like a currency. There is the dollar, euro... and then there are tokens from OpenAI, from Amazon, etc. But in practice, tokens that are priced in dollars and hence rely on English to operate, all end up being cheaper. In other words, when we externalize part of our thinking, metaphorically, by relying on tools like OpenAI, we end up paying in dollars. Which means that, when the dollar rises, we end up paying more.

Indeed, when you use systems based on English, as is the case with tokens, the cost becomes higher as soon as you use another language, such as French. It takes more tokens to express the same ideas in French than in English, a bit like a foreign exchange rate. The difference remains relatively small between English and French, but it becomes much more pronounced for other languages, such as Greek, for example, which consumes even more tokens, and leads to a form of impoverishment.

All of this reflects the issue of sovereignty, particularly when using tools developed by others. From a pragmatic standpoint, it is as if I were plugging my house into my neighbors' electricity grid—with whom I currently get along well. If one day those relations deteriorate, I will find myself powerless. Today, with the major political shifts looming across the Atlantic, this question of sovereignty is becoming increasingly urgent.

Harold ÉPINEUSE: Thank you very much, Maxime, for these ongoing but already quite well-structured thoughts, and for this anthropological perspective that will now be completed by Antoinette's contribution.

Antoinette ROUVROY: To respond very quickly to this video clip, I would say that what we are seeing here is a typical example of the resurrection, under the pretext of technological innovation, of pseudosciences that have been historically discredited. But pseudoscience, even when embodied in new technological devices, remains pseudoscience. When it is used to serve such harmful political agendas, it is surprising that those who engage in it are given any credibility.

To appreciate the societal and political impact of AI, it is worth remembering what Gilles Deleuze so aptly wrote in his visionary “*Post-scriptum sur les sociétés de contrôle*” (*Postscript on Control Societies*) (1990): each type of society can easily be associated with a certain type of machine. Not because the machine determines society, but because it expresses the social forms capable of giving rise to them and making use of them.

The question, for me, is therefore less about how we are going to adapt to these new technologies than about understanding, first and foremost, what social forms are expressed and served by our enthusiasm for devices that behave like machines that metabolize contingency and transform it into rankings, pairings, ratings (scores), profiles, etc. replacing the detection of digital signals and the calculation of abstract, impersonal correlations with investigation, testing, examination, sometimes court appearances, testimony, confessions, etc. What are the social forms promoted by our enthusiasm for calculating the world rather than representing it?

Why, then, are we witnessing this extraordinary craze for this algorithmic shift? In my view, our era is characterized by two major elements.

Firstly, as Luigi Pellizzoni has well described, our era is characterized by an acute awareness of instability, unpredictability, and complexity. Whether in climatology, geopolitics, biology, physics, politics, etc., we have come to realize that we can no longer maintain a relationship with the world based on linear temporality and the regularity of phenomena. Traditionally, we used to represent the world and predict events, even those least accessible to direct perception, through “conventional abstractions” (such as statistical categories) based on the observation of regularities, reductionist equations (which is what probability calculations consist of), which we expected not to be a faithful reproduction of the world, but an approximation pursuing a goal of neutrality. These statistical representations and

probability calculations are not a "second skin" for things; they do not fit perfectly with reality.

But it is precisely this gap—this space between reality and its modeling—that made discussion and criticism possible. This gap implies constant discussion about categories, about which variables are relevant, and which are not, and therefore about how we construct our representation of the world.

We no longer want any of this; categories, discussions about categories. Why? Because the categories through which we used to represent the world now seem insufficiently predictive, ineffective at predicting what presents itself in the form of pure chaos, of indeterministic chaos.

Thus, this increasingly acute awareness of the instability and invincible complexity of the biogeophysical world and, more broadly, of our environments, is linked to a certain faith in the ability of learning machines to magically metabolize complexity in algorithmic black boxes. There is a genuine belief in this magic; we find ourselves at the heart of the utopia or myth of the machine that is smarter than us. It is not a question of comparing ourselves to it; that is not the issue. What matters is its ability to process an enormous amount of data in real time, without producing any "knowledge"—to metabolize this data extremely quickly, abolishing any notion of duration and historicity.

In doing so, "data" is treated as "un-historicized" "facts"; we forget that it is nothing more than the "effects" of events, power balance, and domination. Just as archives can sometimes be nothing more than the frozen history of the victors, machines metabolize data as neutral, ahistorical elements. Through this algorithmic processing, causes and effects gradually become indistinct, creating the impression of greater axiological neutrality. Specifically, the effects in learning machines—including language models (LLMs)—are ultimately nothing more than languages turned in on themselves, encountering only their own effects. These effects are then digested in recursive loops where, gradually, the effects become the causes. This is what complexity of time means: we move from a linear temporality to a complex temporality in which the future determines the past.

And this, it must be said, is something that is profoundly difficult for human beings to assimilate. Yet we trust these machines that operate on a non-linear temporality, one that cannot

be translated into a linear, linguistic form that is intelligible.

Action itself is being transformed: it becomes a form of pure agility. It is no longer a question of predicting, or even preventing or anticipating, but of having the capacity, the agility, to metabolize, in real time, the signals coming from a connected world. Not to "govern" the world, but to "un-govern" it. More precisely, this leads to a world governed by data, that is, a state of facts passively recorded in digital form. Nothing could be less revolutionary and more conservative than that.

This (un-)government by data corresponds to the myth of total immanence, a myth of the abolition of politics, or what Jacques Rancière called political aesthetics. In his book *'Le Partage du sensible'*, the philosopher writes that the fundamental function of institutions is to distinguish between what is a signal and what constitutes noise. For example, in ancient cities, the chatter of women would be considered noise, while the speeches of men, notables and non-slaves, would be perceived as signals, as meaningful. Obviously, this is very normative, dogmatic, and unequal, and we no longer want it. Similarly, the tyranny of language, of the signifier, has a bad reputation today. The French language, which we share—more or less well, being Belgian myself—is stained with colonialism. And there is, in fact, a movement toward emancipation from representation, language, and the idea that data speaks for itself. What emancipation! Would we not finally be guaranteed immediate access, emancipated from the weight of representation to a techno-environmental production of meaning? Immediate access that would even deliver us from anthropocentrism, the absolute evil of our time, to the point of freeing us from the human subject itself.

But then, where does that lead us? I return to the film excerpt: what are we immersed in? Once any "non-signifying" signal can be captured as a digital signal, it no longer functions as a signifier in relation to a signified. It simply functions as a pure calculable signal, even more easily because it travels lightly, being emancipated from any meaning. In other words, it is freed from any relationship between thought and the world. What matters, as in the linguistic turn, is simply the purely superficial relationship between data and other data. The material world, empirical reality, is forgotten. What does this mean in concrete terms?

Practically, this would mean that we live in a deeply paranoid world in which any signal can enter a calculation process designed to detect everything and its opposite. What these algorithms essentially produce are speculative spaces and a way of acting that could be described as “un-futurizing”. These are conjectural techniques that allow us to act immediately, in an ecosystem of ignorance, since we no longer understand the causes of phenomena. Nevertheless, we respond immediately to alerts about what still exists only in a world of potentiality. This is a completely new mode of action, but one that is eminently paranoid (well-illustrated by the concept of the “operational image” in Harun Faroki’s cinematographic work).



Practically, this would mean that we live in a deeply paranoid world in which any signal can enter a calculation process designed to detect everything and its opposite. What these algorithms essentially produce are speculative spaces and a way of acting that could be described as “un-futurizing”.

There is an obvious appeal in this idea of having a supposedly immanent relationship with the world, but that is completely misleading. Why? Data are not simply given; they are always produced in a certain way. Even what we call raw data does not exist in nature as such. It is always the result of a long and patient, rather sophisticated process of anonymization and deindexing. In other words, we must organize the amnesia of data in relation to its conditions of production, which represents an opportunity for depoliticization. And behind this issue of depoliticization is the desire to give the impression of a kind of neutrality in an immanent space.

However, this metric coordinate space, whose zero point is no longer the human being, corresponds more to this idea of a kind of speculative realism, or what I call algorithmic realism. Today, we are seeing the emergence of this technical

ideology of big data or digital behaviorism. And, ultimately, all these elements make us forget that behind this façade, there are neo-institutions, which I call “hyper agents”, the most typical examples of which are probably Elon Musk or Mark Zuckerberg. It is the hyper agents who, ultimately, shape the algorithms.

An algorithm that improves itself through machine learning fragments the world into meaningless, infra-personal data or signals, and then recomposes it with a kind of automatic curiosity on the part of the machine, domesticated by the objective function of the algorithm. The latter corresponds to the formalization of the sectoral logic or interest of the hyper agent who uses the algorithm to shape the world in his own image. He will maximize or minimize certain things, ensure that some inevitably happens, according to his own interests.

It is therefore not a smooth, neutral space at all, but an eminently striped one. Within this space, it is no longer norms that are produced, but a kind of lawless force of these new hyper agents. Thanks to this increased agility, they have made themselves capable of metabolizing what is ungovernable, that is, this indeterminacy, and of maximizing contingency, intensifying volatility to their advantage. Nowadays, when Silicon Valley players are literally storming governments, what we are ultimately witnessing is the integration of speculation into the very heart of governmental rationalities.

It is a joint tour de force of these shifts in statistical rationality and neoliberalism, which has also placed speculation at the heart of governmental rationalities. What can we do about this?

First, in my opinion, the issues at stake are not those perceived today. There is currently a kind of obsession with AI regulation and ethics. I probably will not please everyone by saying this, but I do not understand this idea that machines need to be made ethical, and this rise in the importance of ethics, which is ultimately becoming the dominant rhetoric in discussions about innovation governance.

It is a form of techno-reformism. To make machines perfect, we decided to create fair, accountable, and transparent algorithms, as if justice could be a property of a technical subsystem, of a non-algorithmic box designed abstractly from its uses. This is a complete denial of the materiality of the world. We do not know exactly what justice is.

Justice presupposes that we catalog injustice, because it is first and foremost a concept whose truth lies in dialectics. Justice is a fundamentally dialectical concept, which is why it is never present and cannot be found in data. It represents something to come and, as such, must remain open and indeterminate. On the other hand, what is visible, what can be described, is injustice, and it is precisely reflected in the data.

Data is the effect of the largely unsustainable and unjust state of things, in the present. So, this is what needs to be made visible: the injustice and the shift from a regime of normativity to a regime of facticity that rigidifies history into destiny, a common feature shared by algorithmic regimes and fascist regimes. When history is rigidified into destiny, we ensure that the angel of history, as evoked by Walter Benjamin, has no space left to arrive and correct history. Data—or at least the facts that gave rise to this data—since they are the effects of power relations, must be corrected to do justice to the real situation of individuals and their singularity. Not to the particularities detectable in large numbers, but to the singularities of situations.

The first thing to do, therefore, is to do justice to the facts and not to the data that are their effects. Secondly, we must recognize that the issues at stake are perhaps less issues of regulation than issues of a constitutional nature.

It is in the constitution that collective criteria for merit, need, desirability, and acceptable danger are decided, which govern the distribution of opportunities in social life. It is not acceptable to outsource to machines the responsibility of determining for us, within the mass of data, what these criteria are. It is essential to decide together on the theory of justice that concerns us.

Furthermore, among the virtues of a constitutional text is this: it does not impose so much domination as it limits the powers of those who hold them. It is a text that limits power. It so happens that today, we must limit the powers of these “hyper agents” I mentioned earlier. These new institutions act in a true Wild West fashion and confer on themselves a lawless power, or lawless force, which must be tempered. We must think of constitutional architecture that constrains them.

Another virtue of the constitutional text is its temporal virtue. The constitution is also the manifestation of a capacity or technique of temporalization that is characteristic of law. In other words,

it is a commitment that transcends the very life or presence of those who sign it. This also allows us to consider those who have not yet left any digital trace and future generations, left behind by these “hyper agents” who fantasize about departures to Mars, which will probably not concern anyone.

This would also impose limits on extractivism and the depletion of resources, to which we are condemned by these massive investments in the development of artificial intelligence. The training and use of AI, particularly generative AI, is, in the literal sense, exhausting in terms of energy resources and the climate. And if there is an existential risk, it would likely lie in this area. Personally, I do not believe in the risk of hyper-artificial intelligence eventually ruling us. On the other hand, resource depletion is something tangible and something we absolutely must be worried about.

Harold ÉPINEUSE: Thank you very much, Antoinette, for this political perspective, which also brings up economic, geopolitical, and environmental questions, and these endless promises of AI in a finite world, at least for now and probably for a long time. What about the sociology of science? Bilel, you have the floor for this other perspective on societal and political impacts of AI.

Bilel BENBOUZID: As a sociologist, my style of discourse will be a bit different, maybe a bit more down-to-earth. Before making the connection with the documentary, I would like to share with you a small survey that we have just started it on the emergence of assistants—particularly ChatGPT—in the realm of education, at high school and university levels.

To begin and contextualize this survey around the major issues of AI regulation, here are some excerpts from interviews. A high school student, for example, told us: “It is either ChatGPT, or I will get a bad grade”. Another one, and we are talking about rascal high school students who are struggling, told us: “I have always trusted it, it is always relevant, and my good grades thanks to it confirm that it works well”. And here, you have a student who is doing well presently. What will happen with the final exams, we do not know yet, but for now, he is doing quite well.

And then, another type of profile, a third high school student, quite serious, who had always tried not to use it. The researcher asks her why

she started using it since she had not done so until now. She answered that she was afraid of relying too much on this tool, later. According to her, many people installed it thinking it would simply be a help, and she continues by saying that afterward, she saw them cheating with it during classes and doing all their homework with it, and that they ended up relying on it all the time. But in the end, she says she feels a bit forced to use it because she is in a competition with her classmates regarding grades, and using AI, sometimes even the worst students become stronger than her by cheating.

It is interesting to see how this technology has entered the lives of these high school students, while nothing has overseen the arrival of these machines. Law and regulation have radically failed.

At the university level, you can find a student who is writing her thesis. She confides that she uses it intensively, and that it serves as emotional support as she feels less alone in her work. She explains that in the paid version, there is a memory function, which is very practical. For a long-term project, it remembers all the discussions and exchanges, and gives personalized advice. Moreover, depending on the tone used, it will respond in the same tone, which she finds funny and makes her laugh, as ChatGPT uses friendly expressions like "Hang in there, you will make it!" or "We are going to wrap up this thesis together, we are a team!". And this student feels comforted by that, even though she knows it is only technology. So, she feels less alone, knowing that she can delegate tasks when she herself no longer has the energy to work some evenings.

As a sociologist and anthropologist, I am fascinated by this ongoing transformation of the world, and on which we have no political control. So, we will conduct surveys, as always, on the usage and practices. There have already been surveys on social networks, on the internet, and each time it generates debates. These surveys are part of a larger reflection on the sociology of science and how to govern science and technology. And here, you find Collingridge's dilemma. Indeed, the problem with science is the lack of initial information. It is necessary to set them up to be able to observe their impact.

Right now, we are in the process of collecting this data, but it is only once we measure it, in a few years, that we will see the effects on students. By then, these technologies will be

well-established, and we will have become dependent on them, as they evolve alongside us. And this represents a dilemma.

We could react to this by saying that we will engage in participatory democracy and define these AIs as we want them to be—the famous technical democracy. We might imagine a kind of constructivism in the vein of Bruno Latour, even though today, the "Latourians"—of whom I was once part, being trained as an engineer—are somewhat outraged by the world we live in. But it is a bit late, after 30 years of sociology of science that have allowed many things to slip through in terms of innovation.

The reason I bring this up is because of a compass that we created with Yannick MENECEUR and Nathalie Smuha. When I started working on AI five years ago, I long believed that by looking at the uses, we would finally be able to map the controversies, and once this map was drawn, we could perhaps organize a public debate and build the AI we desire for our societies.

Initially, I still believed that, and we had made this compass that I am now changing. There is the neutral axis, anti-neutral, which we had not called this way at the time, and I am in the process of changing it. In fact, it is like Magritte's painting, "this is not a pipe". This is not a controversy, whereas at first, I thought it was. It is ambiguous.

Jean-Gabriel Ganascia told us how, in the end, companies like IBM changed sides. As a political scientist, among other things, I would like to take these ambivalences, and what I have called a political blur, seriously. We could even speak, in the image of Philippe Corcuff, of "confusionism", confusion. Elon Musk, for instance, a friend of Barack Obama, is now a friend of Donald Trump. Mark Zuckerberg, who had funded research against mass discrimination for Meta, is now bringing back masculinist energy to the forefront. What can we make of this blur? Either we can simply state that we are being manipulated, that we are being mocked, or we can, in the manner of Philippe Corcuff, try to understand what is structural in this confusionism—or blur—and what functions it is serving. The word "function" here being understood in the sense of Roger Girard, meaning the function of a structure in the public space.

How did we build the compass? We started from the assumption that we do not know what AI is, as Antoinette said, it is a figure of speech. We

can say that it is both something quite vague but also talk about its concrete effects and its progress, for example, when talking about ChatGPT. The word is both abstract but has a concrete meaning. When we can use a word in the same sentence in both ways, it is called a syllepsis in rhetoric. It has a strong effect; it is something we project into the future, very concrete, but at the same time, we do not really know what to make out of it.

Thus, we created an axis, the axis of rhetoric, which conveys something either very concrete or abstract. A second axis had been drawn, which I have now modified, and I call it political, neutral, and anti-neutral. As an engineer, and a good reader of Langdon Winner, for me, technology represents the values embedded when we created this compass, and it is also what it imposes when we must make it turn. This corresponds to the political meaning of techniques. And I somewhat criticized engineers who thought technology was neutral. Over time, however, I began to take the issue of the neutrality of techniques more seriously, which is quite troubling, notably thanks to Roland Barthes' work "Le Neutre" (*The Neutral*).

I have been teaching in engineering schools for a long time, and I must admit that they can be stubborn and think that talking politically about technology is just nonsense, literally. They had drawn a pipe, the music instrument, on the school newspaper next to my sociology of science course with the inscription, "If you go to Mr. BENBOUZID's sociology of science course, cut out this pipe". They did not like that class—even though I did not always spare them as well, I must admit—because they do not understand the political importance of technology. It is a Pareto curve, 80% of them think it is nonsense, and 20% become sociologists—which is already a pretty good rate, of course. I am telling you all this because we can observe these same engineers in the controversy.

There are thus four spaces within this controversy. We can find the neutrals, those who believe that technology is neutral, and the anti-neutrals, those who believe it is not. But ultimately, even when considering this desire for neutrality among engineers, as Roland Barthes would say, the neutral is not lukewarm; it is scandalous, and in the end, it is already a stance. Still, technology cannot be placed on the left or the right, as illustrated by Musk's shift from Obama to Trump. This

neutral/anti-neutral divide is possible. And within it, we can project various actors.

Among them, the transhumanists, the arena number one, who are in fact anti-neutral, which is quite troubling. Secondly, in my field, STS (Science and Technology Studies), we find a few engineers who consider AI as a sociotechnical system. In contrast, the third space includes more traditional engineers who do not align with the transhumanists and strongly disagree with their discourse. These are typically academic AI researchers, who view the issue as one of safety that can be addressed by developing standards. Finally, there is European regulation and the law, which surprisingly has become nearly an actor in its own, right among all the others.

And what I want to show is just how confusing this whole landscape is. For example, Jean-Gabriel Ganascia mentioned the transhumanists' moratorium. We did not understand why the very actors themselves called for a halt in the production of generative AIs. Another source of confusion before that was the dismissal of Sam Altman. It might be framed as a kind of effective altruism versus effective accelerationism, a new category altogether. What emerges from all this is a sense of various ideologies, though truth be told, there are not any. It is a kind of a blurring, confusing place.

Here, you can witness female researchers who have achieved extraordinary things and for whom I have great admiration. They did not hesitate to mock ChatGPT in a highly critical article where they referred to it as a "stochastic parrot"—a phrase that has been widely echoed, even if, in my opinion, it is a bit more than that. However, what struck me and what I appreciated in their critique is that they finally set boundaries, something Google refused to do.

Nevertheless, something that bothered me, and this is why I speak of confusionism, is that for a long time, these same researchers, including Timnit Gebru, called for fewer men in AI. Of course, addressing issues of discrimination by encouraging women to enter the AI field is important. But would that solve the fundamental issues related to the politicization of AI? Because in that case, we would also have to consider Queer AI, Black AI, and other such identities in response to different forms of discrimination. And even though I deeply admire these individuals, I am not sure if that would truly be beneficial for AI regulation.

The worst part is that research, in general, was massively funded by the very same actors who have now aligned themselves with Trump, further blurring the compass. Yann Le Cun, among others, a staunch critic of transhumanists, constantly repeats that AI will save lives. In doing so, he too adds to the confusion: how exactly would AI save lives? On the contrary, we are witnessing a questioning of the safety culture among engineers. Self-driving cars are put on the market based on their capabilities and performance, even though they are not safe.

We are seeing a transformation in risk management paradigms. ChatGPT was placed in the hands of students without anyone knowing its effects. Contrary to what Yann Le Cun claims, we are far from saving lives; instead, we are transforming our relationship to risk. Although machines may be competent and highly performing, they are not necessarily reliable, and this new politics of reliability seems deeply troubling to me.



Contrary to what Yann Le Cun claims, we are far from saving lives; instead, we are transforming our relationship to risk. Although machines may be competent and highly performing, they are not necessarily reliable, and this new politics of reliability seems deeply troubling to me.

Finally, the height of the confusion: regulation and the AI Act. This regulation was supposed to create fundamental rights—but it does not; to manage risks—but it does not do that either; and to contribute to market construction. Yet the market is being built elsewhere than in Europe. In truth, we are in the middle of the fog, pretending to be able to control the flow of events.

It is precisely this problem of confusionism that lies at the heart of the rise of far-right movements today. Philippe Corcuff has analyzed this mechanism by which a discourse always borrows from its opponent, but for its own benefit. Thus,

the far right borrows from the left, while constantly covering their tracks. And it is somewhat the same here. We wanted to reassure each other by drawing on decolonial movements, while at the same time being transhumanist, with traditional engineers focused on safety, all wrapped in a regulatory framework. And the worst part is that within this regulatory framework, the decision was made to put open source at the center of the system. Open source is a kind of regulatory sandbox, which will also become a source of radical constraints.

So, what is the future of AI? In the years to come, we will have no political control over this technology. And paradoxically, it feels like we have never debated it more, never wanted to control it more. It is a rather pessimistic observation, but I see no way out. That is where our political problem lies.

Harold ÉPINEUSE: Thank you very much, Bilel, for shedding some light on this thickening fog, helping us to find our way—coming from an engineer who turned to sociology. I would also like to take this opportunity to emphasize that at the IERDJ, our work in the humanities and social sciences is, of course, open to collaboration with engineers. We are fully aware that all stakeholders must be at the table, as is the case today, to work on these issues. And finally, to conclude this multidisciplinary perspective, we turn to a philosophical and legal practitioner's approach. Sabine, the floor is yours.

Sabine VAN HAECKE LEPIC: I do not know if I will be able to "un-blur" things for you, but in any case, I will try to address the topic. Will artificial intelligence transform freedom into destiny? Through this graphic revolution and these meaningful tokens, it is our very language that is being altered, giving rise to a new meaning in writing. A new language is emerging—a mathematical language made of words. As Professor Ganascia put it, it is a form of distributional semantics. It seems to make sense, but has it not somehow captured or overtaken the spirit of languages? In any case, as Murielle said, words travel, calculations disguise themselves as words, and the confusion is complete.

In this graphic revolution, the question arises of decision-support systems built on probabilistic models, which raise significant philosophical issues and have major societal impacts. With

some of us keeping daily watch over these new brains and cognitive prostheses that AI systems represent, the question of how our freedom is being transformed becomes indeed relevant.

How does AI transform our freedom into destiny? Freedom is a fundamental issue, and new technologies offer an opportunity to reflect on the future of our humanity. What do we want? Are we truly free? Does the meaning of this question take on its full depth when tested against these AI systems? What kind of society do we want?

For this sequence, I have drawn on past and contemporary philosophers who, for several centuries, have been reflecting on the recurring question of our freedoms and the reality of our relationship to technology. Let us decode with them what is at stake and the question of possibility and freedom.

Regarding the possible and the real, I first turned to Bergson and his 1934 book "La pensée et le Mouvant" (*The Creative Mind*). The philosopher, vitalist of duration, intuition, and change, develops his thoughts on possibility. He writes:

"I would like to return to a subject I have already discussed, the continuous creation of unforeseeable novelty that seems to go on in the universe. For my part, I believe I experience it at every moment. No matter how much I try to picture in detail what is about to happen to me, how poor, abstract, and schematic my representation is compared to the event that is taking place!"

In the end, this is somewhat what is happening today. He continues:

"The realization brings with it an unpredictable nothing that changes everything. For example, I must attend a meeting; I know which people I will find there, around which table, in what order, to discuss which issue. But whether they come, sit, and talk exactly as I expected, saying what I thought they would say, the whole experience gives me a unique and new impression, as if it were now drawn in one original stroke by an artist's hand. Goodbye to the image I had made of it, a simple juxtaposition, something that could have been figured out in advance, of already known things!"

In this excerpt, Bergson refers to a superhuman intelligence that could predict the future state of the universe if it had all the information, and I would say today, all the data. But the philosopher reconsiders and contradicts himself, insisting that the organic, the living, the mind are not lifeless matter. And even more, they are not

reducible to their predictions. However, today, this attempt at algorithmic determinism has taken over.

Are we heading toward the programmed obsolescence of reality, of the improbable, and of the realm of possibilities? The debate on the predictability of phenomena is not new. The materialist determinism recalls Laplace's demon in his "Essai philosophique sur les probabilités" (*Philosophical Essay on Probabilities*) from 1814. An intelligence with absolute analytical capabilities, possessing all the data of the universe, could predict everything. Nothing would be uncertain for this form of intelligence.

Kant expresses this in its "La Critique de la raison pure" (*Critique of Pure Reason*) in 1781. Causality according to the laws of nature is not the only one from which all the phenomena of the world can be derived. It is still necessary to admit a free causality for the explanation of these phenomena. This statement is decisive as it separates the absolute determinism of Spinoza from the autonomy of the subject in Kant.

These elements, before beginning a process of reflection, must be placed in the perspective of a generative artificial intelligence that eliminates singularities, amplifies biases, and determines us. Am I deciding myself, or am I guided by algorithms? Am I free? At this moment, it is about confronting my desires with the recommendations. You said destiny, a closed and deterministic circuit would be set up.

Yes, artificial intelligence transforms our freedom into destiny. Every algorithm based on neural networks for natural language processing, or NLP, is built to predict the next sequence, just like autocompletion software. These are present on our phones and suggest completions for our messages. The main generative artificial intelligence models, similarly, are based on statistical calculations aimed at producing the most probable sequences based on prompts. These models are therefore average standards and are incapable of producing original, improbable, or unique content.

This is the language of the probable, as Yannick has demonstrated. The consequence of this is that it amplifies the prevailing trends and eliminates the possible in favor of the probable. Artificial intelligence thus generates standardized content, but it does not stop there. These probable contents will then be trained on content that has already been produced.



The main generative artificial intelligence models, similarly, are based on statistical calculations aimed at producing the most probable sequences based on prompts. These models are therefore average standards and are incapable of producing original, improbable, or unique content.

We would be witnessing, as subtly suggested by Antoinette, a kind of closed loop, or even auto-phagy of the artificial intelligence thus created, with serious consequences: homogenization, uniformity of responses, questioning of active, improbable, and singular content, the risk of synthetic data mixed with real data... A generative algorithm in motion that would reinforce determinism and give a probabilistic prediction to foresee with certainty what would happen to us.

There would inevitably be a reduction in the field of possibilities, with little or almost no social mobility. I did not react to the film, but it is an absolute determinism on morphology: no atypical paths. Some would be quite troubled here. A reduced field of possibilities, nourished by algorithms and probabilities.

Liberty, destiny, determinism. This questioning touches on the royal prerogative, the function of judgment. If we are heading towards technological solutionism in the public sector, a double set of questions arise. The decision-making approach through a decision tree to assess the necessary use of generative artificial intelligence systems in the public sector. In the case of a favorable decision—I quote Yannick—it is necessary to resort to democratic oversight. But regarding this determinism, the question brings forward other specificities when these systems are used within the public sector, with a risk of a governing machine.

And here, we move to December 28, 1948, with the birth of a new science, cybernetics—towards Dubarle's machine to govern, updated by Arnaud Billon. Will the mechanical manipulation of human reactions one day create the best of worlds? Mass statistics, the first great relays of the human brain.

All these machines share a common characteristic. They collect data, work methodically based on a specific problem they are to solve, more or less perfectly, using suitable data, until they provide a solution and execute, if necessary, what this solution prescribes. Moreover, they do this work better, in a much more complete manner, and especially much faster than humans, left to their usual capacities. Let us say that they are machines designed to gather and elaborate information to achieve results that can be either decision or knowledge outcomes.

These creations were still in a phase of evolution at that time, but the first stirrings of this new technique already attested to a clear surpassing of the organic powers of the human brain. It was only 1948, at the dawn of cybernetics, and yet the impression arises that we are speaking about it as if it were 2025, questioning its consequences, the potential paralysis of consciousness, and the emergence of a prodigious political leviathan in the vein of Hobbes. Today, we risk a massive global city where primitive injustice, deliberate, and self-aware, would be the only possible condition for a statistical happiness of the masses.

Let us come back to 2025 and the questions that we must ask to avoid this technological solutionism. Ethical questions, but not only that. We need to identify a criterion that is not misleading in interpreting machine decisions. Otherwise, when decisions directly affect individuals, dehumanization is looming with total technological solutionism, once again, with the current parameters and deployed functionalities.

It seems difficult to understand the decision-making process. Therefore, to address this inadequacy, it seems necessary to intensify the need. How can we do this? Is there a solution, or a diversity of alternative solutions depending on different economic models? Earlier, we were wondering if there was a ceiling, but in my view, it is more of a wall: the wall of our energy cost of using AI, but above all, the wall of global resources. This wall is not good news, but at least, it may force us to rethink things differently because the choice is no longer allowed.



Earlier, we were wondering if there was a ceiling, but in my view, it is more of a wall: the wall of our energy cost of using AI, but above all, the wall of global resources. This wall is not good news, but at least, it may force us to rethink things differently because the choice is no longer allowed.

It may therefore be possible to extract good from bad. If we had no other choice but to turn to more frugal technologies, perhaps models that operate closer to the user, a desirable geopolitical reflection would emerge. There could be a focus on an economic performance tool closer to the uses, and a resetting of territories closer to the user, with somewhat more closed circuits and less risk of backdoors. Wishful thinking, I do not know.

Are we going to make generative AI disappear, as it is currently practiced? That is Luc Julia's opinion, who is currently deconstructing the myth of artificial intelligence. Would it not be, after all, good news that this very system, which makes us addicted and directly influences our psyche from a human ecological point of view, is also not viable in terms of ecology in the broader sense? In other words, this capitalist system is based on capturing our data, which requires ever more powerful data centers, etc. Maybe this is an opportunity to think differently and move towards a more sustainable, more human ecological reflection, to break this determinism.

Let us invite Darwin into our reflections, to our rescue. What may be at stake is an evolutionary selection process of the most performant, allowing the elimination of others, particularly those who cannot anticipate the law of probability of the world. The day our world will no longer be populated by anything but disembodied beings, producers of rational anticipation, freed

from everything that still prevents us from perfectly calculating the law of reality. On that day, a bifurcation will take place, markets, having become absolutely and definitively efficient, will take power, and humans will finally conform to the *homo economicus* of neoclassical theory.

It will no longer be the elimination of humans surpassed by super-intelligent machines, marking the end of history, but the elimination of the less performant, ruined by the law of supply and demand. So, if ideologues, to preserve the spirit, advocated for a disembodied individual, with the triumph of agents, all would together pursue the project of liquidating collective institutions and singularities.

Destiny or the freedom to conceive the scenario otherwise. The new technologies, most of whose benefits will turn into as many threats, need to become the subject of a political debate to become contributive and hermeneutic digital commons, and not hegemonic and disruptive innovations that bypass public power, to quote Anne Alombert. If the public institutions inherited from the last century are neither equipped nor sufficient to face such challenges, their pure and simple elimination through libertarian logics will only lead to their liquidation. If institutions have no choice in the face of these major transformation challenges, it is imperative, as we are doing today, to come together with such intelligence to respond to these challenges and to face the machine, to create a real opportunity for justice and for our democracies, which are so fragilized nowadays.

Their transformation and reinvention are what is needed today. These do not depend on the computational performance of automated machines recombining the probable data of the past but require today a colossal courage and effort of collective reflection and deliberation, opening to the improbability of a truly intelligent future. I will end with a quote from Nietzsche that I particularly appreciate: "It is not doubt but certainty that drives you mad...".

Harold ÉPINEUSE: Thank you, Sabine, for bringing together these big names in philosophy to enlighten us and for this political perspective on AI.

Maxime DERIAN: To quickly follow up on Antoinette's remarks and the hyper agents, in my concept still in progress, I talk about the "hive's

queen". These are individuals who share the same DNA but are nurtured differently and play a more specific role. This role is amplified by hybrid agents and AI that relay their decisions. Tolstoy, in "War and Peace", asked whether Napoleon himself decided to attack Russia (or invade Belgium...), or whether the people had driven an impulse in that direction.

Are these idols really the leaders? When we see Elon Musk tweeting 130 times a day, we are starting to know everything about his daily life. If the emperor is naked, we can ask ourselves where he is leading us. Certainly, he is not going to lead us to invade Belgium or Russia, but I was thinking of this quote from Nietzsche: "No shepherd and one herd! Everybody wants the same, everybody is the same: whoever feels different goes voluntarily into a madhouse."

Antoinette ROUVROY: It is a convergence of highly disparate elements. What strikes me is not the homogenization effect, it is rather quite the

opposite. They lead their flock and attract, in a magnetic way, very diverse individuals who often have little in common. Voters who had no real interest in seeing Trump come to power voted for him—and the same goes for Elon Musk's followers. They tap into extremely powerful mechanisms, not so much of seduction, but of resentment. People are isolated, and through resentment—finding scapegoats to hate together—they seek new forms of affiliation or allegiance. This phenomenon is not uniquely American; it is spreading across Europe as well. Hence, the importance of rethinking how to produce more desirable commons, to imagine new shared social and political consistencies.

Harold ÉPINEUSE: Thanks to the four of you for feeding our reflection and for all these references. This very rich material, which gave us much to think about, will allow us to gather around the concrete operational issues of AI, particularly in the legal and judicial professions. ■

Round table: AI and legal professionals, a great misunderstanding?

- Deciphering expectations and issues linked to the use of AI in the law: greater efficiency, to be objectivized
- Audacity, transparency and caution appear central in the use of artificial intelligence systems in the legal field
- Concrete projects already in the legal professions

Moderation: **Sophie SONTAG KOENIG**

Senior lecturer at the University of Paris-Nanterre

Aurélie KLEIN

Lawyer and Associate Lecturer at the University of Strasbourg's Faculty of Law, Political Science and Management

Frédérique BERROD

Professor of EU law at Sciences Po Strasbourg

Pascal ALIX

Lawyer

Claire STRUGALA

Judicial magistrate and Doctor of law

Marc CLÉMENT

Administrative judge

Harold ÉPINEUSE: We will resume the course of our program with a panel moderated by Sophie Sontag Koenig, a lecturer at the University of Paris-Nanterre. We have long-standing ties between our institute and Sophie, as her thesis won the Vendôme Prize, an award for young research granted in partnership with the Ministry of Justice and the Directorate of Criminal Affairs and Pardons. She recently published a summary article, providing an overview of audiovisual devices for legal and judicial information, particularly focusing on online dispute resolution, which was published in a collection titled "État des connaissances".

In the same collection, you can find the article by Murielle POPA-FABRE on generative AI and judicial decisions. Additionally, the proceedings

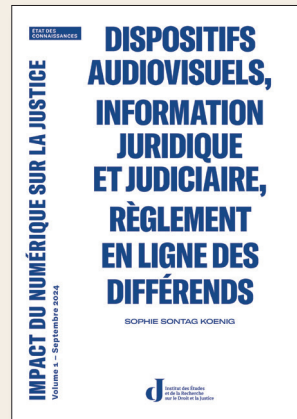
of the Cyberjustice conferences, organized in partnership with the Council of Europe and supported by the University of Strasbourg, are available online, with the latest volume being on the new challenges of cyberjustice, between information, regulation, and democracy.

Lastly, I would like to pay tribute to Yannick, who designed this day for all of us, and to his book published by LexisNexis, "IA générative et professionnels du droit : comprendre et s'appropriier la langue des probables" (*Generative AI and Legal Professionals: Understanding and Adapting to the Language of Probabilities*), which is already receiving well-deserved success. With these words, I will hand the floor to Sophie for this second round table.

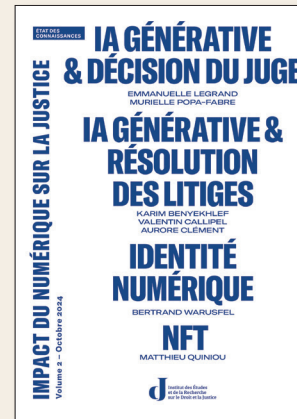
Sophie SONTAG KOENIG: First, I would like to address a few words of thanks to you and Yannick for your renewed trust, for allowing me to lead this roundtable on a rich and fundamental topic for the many students that came today and the professionals we are going to represent. Thanks also to the IERDJ, who make these exchanges possible, and to the University of Strasbourg for their warm welcome. Standing in a university to discuss these professional issues is both symbolic and valuable. The mission entrusted to me now is to "grasp AI" as "to leap into the work of legal professionals", as it was said this morning.



<http://bit.ly/4hblyfz>



<http://bit.ly/3KSBKA7>



<http://bit.ly/47fByCw>

As opening remarks, on December 18th, 2024, the French Senate published a report on artificial intelligence and legal professions. The rapporteurs highlighted several points, especially the gap that can sometimes exist between the high expectations that legal professionals are starting to have regarding AI. However, as has been pointed out several times, AI is a term whose boundaries are not always clearly defined. A technological gap is thus emerging and widening between certain professions. This is why I am pleased to moderate this discussion with several professionals who will share their complementary analysis on these issues. Together, we will try to understand the challenges that AI brings to their profession and how it interacts with others.

I will briefly introduce them, despite their many specializations. Aurélie KLEIN is a lawyer, expert in data and digital at the FIDAL firm, and an associate lecturer at the Faculty of Law at the University of Strasbourg. Frédérique BERROD is a professor at Sciences Po Strasbourg. Pascal ALIX is a lawyer, specializing in artificial intelligence law and data protection, and a PhD candidate in private law at Paris 1 Panthéon-Sorbonne. Claire STRUGALA is a judicial magistrate and holds a doctorate in law. Lastly, Marc CLÉMENT is an administrative judge and holds other more technical roles, which he will clarify for us. Lawyers, judge, professors and researchers—while not all legal professions are represented—we already have a rich panel that promises complementary discussions.

We have divided our discussions into several axes in which each of you will have the opportunity to participate and respond to others'

thoughts. However, to begin, since AI is such a broad concept, it is necessary to clarify the scope we will assign to it in this round table. Generative AI has been widely discussed so far, but AI does not stop there. So, what is the scope of AI that you feel is most relevant to discuss in this exchange related to legal professions?

Marc CLÉMENT: Hello, I just wanted to clarify, in relation to what was discussed this morning, that I belong to the category of engineers who still do not consider technology to be neutral and who are closely interested in this issue. In the 1990s, I notably served as the deputy director of a computer science laboratory within a large company, where I used artificial intelligence. It was a period still favorable for this field, before the momentum somewhat relapsed.

Afterwards, I joined the administrative courts and am now a first-instance judge. I must precise here that my intervention today is strictly personal, and I am not speaking as a representative of administrative jurisdictions. My experience therefore combines that of a first-instance judge, but also that of a developer. I apply AI techniques on a website called 'www.lawdataworkshop.eu', where, for example, you can find the transformation of administrative court judgments into music.

How, then, could we rightly define the scope of AI? Today's initial discussions have allowed us to see that there is a complex and rich history behind the concept. When we look at the legal definition of artificial intelligence—as formulated, for example, in the European regulation on AI—one cannot help but feel a certain perplexity. In my view, the legal debates around this definition

are not very enlightening. On the contrary, they tend to create confusion and be sources of perplexity.

Indeed, if we take this definition to the letter, it ultimately applies to any program or software. We almost end up encompassing all software. Because as soon as we assert that a program pursues a goal and will develop a certain number of possibilities to achieve it, we are describing what computing has been doing since its beginnings. This is, by definition, what software does.

There is indeed a real hassle in legally apprehending an object in constant evolution. For example, during discussions at the European Parliament around the regulation on artificial intelligence, the issue of generative AI emerged even though it was not part of the European Commission's initial proposal. This illustrates the necessity of addressing new technical objects. This morning, for example, a visual aid presented different aspects related to generative AI, highlighting computational concepts that, although quite technical, allow for an approximate delimitation of the scope within which these technologies are situated. And yet, one can see how difficult it will be to maintain such a map, even over the next two or three years, given how likely it is that future evolutions will be quite radical.

There remains a huge amount of legal definition work to be done on these technological objects. Personally, I tend to think that the central issue lies less in the exact nature of what we call artificial intelligence than in the positioning we take towards algorithmic decisions themselves. In other words, what is fundamental is to take a

stance on the idea that algorithms will intervene. The nature of these algorithms, whether they correspond to AI or not, and their level of intelligence, ultimately have a rather limited impact.

Another essential aspect on which the European regulation on AI seems to advance, without truly stating it, is that of uses. There is no such thing as a technique, in itself. One does not pick a box from an AI vendor's shelf and say, "I have a product called AI."

What is put in place are tools embedded in specific institutional contexts, with concrete users. It is here, in this interplay between technology and use, that real stakes appear. And this, it seems to me, is what deserves our full attention.

Sophie SONTAG KOENIG: Aurélie KLEIN, Pascal ALIX, you are both lawyers. What is your profession's complementary perspective on this matter?

Aurélie KLEIN: To revisit the points already discussed and explain why we are particularly focused on generative AI; it is simply because it is the phenomenon that slapped us in the face about two and a half years ago. It is practically impossible to avoid this topic. One cannot not read articles, not listen to editors, not meet Legal Tech companies that often promise great things, without talking about generative AI. In the face of this massive enthusiasm, we have therefore taken up the subject within my law firm.

I am a lawyer specializing in data law, but I also am the head of digital innovation at the FIDAL law firm. In this framework, we think a lot, observe, and engage in iterative reasoning by conducting many tests. The reason we talk a lot about generative AI is because there is a risk of getting misled into something that might have its limits. We need to avoid "juridication", which is already seen with the Doctissimo website and the healthcare sector. The danger for us lawyers, and it is a real one, is having clients come to us saying, "I asked ChatGPT, here is my case, please draft the document as I dictate.". This is why we wanted to share these points of reflection with you, to avoid these absurd and dangerous situations.

Sophie SONTAG KOENIG: As a matter of fact, we will come back to the risks (it is one of the discussion points we agreed on!). But before that, Pascal ALIX, what is your reaction to these preliminary remarks?



There remains a huge amount of legal definition work to be done on these technological objects. Personally, I tend to think that the central issue lies less in the exact nature of what we call artificial intelligence than in the positioning we take towards algorithmic decisions themselves.



Claire STRUGALA, Sophie SONTAG KOENIG (moderation), Frédérique BERROD, Marc CLÉMENT, Aurélie KLEIN, and Pascal ALIX.

Pascal ALIX: To begin with, what are we talking about? Does AI exist? I have never met it, to echo Antoinette ROUVROY's intervention. And, truthfully, if we look at the regulation, we are caught in a trap. Three institutions—the OECD, the Council of Europe, and the European Union institutions—have agreed to adopt a definition of the concept of artificial intelligence, which is questionable because it is not really one.

The choice could have been made to rather speak about algorithmic decision-support systems. In fact, I co-signed, along with Yannick and others, an article contesting the confusion created by the use of this term, which is not really one of "artificial intelligence". Why this word, "intelligence"? This label is based on an initial anthropomorphism, which misleads us into making a comparison. It is impossible, of course, to compare the embodied human mind with these algorithmic systems. This is a first major source of confusion.

Secondly, regarding the scope, since that was the question asked—my comments so far have been mostly introductory. Concerning the scope, since we are here to talk about our experience, I will speak exclusively about generative artificial intelligence and how it fits into a law firm daily. It seems useful to narrow down our reflection, so

we can then better detail the advantages, limits, and difficulties related to the use of these technologies.

Today, it is clear we are dealing with an object that is both sociotechnical and political. It is not just a tool—contrary to what we may read or hear everywhere—but a system that, potentially, provides services. We could also debate this: does it truly provide services? Is it really something new? These questions deserve to be discussed, starting from our practice, from the concrete.

Sophie SONTAG KOENIG: Perfect remarks to kick off this debate on practices, during which, rather than discussing "needs", we will focus on the "uses" of AI. Claire STRUGALA, as a judge in the judicial system, can you tell us whether AI, generative or otherwise, is already being used? We are gradually coming to understand that, despite a morning dedicated to trying to define it, each person ultimately has their own definition, their own need, and their own use of AI.

Claire STRUGALA: Hello, I am indeed a judicial magistrate and have had the opportunity to follow the dual negotiation at the Council of Europe and the European Union on the AI Act regulation. Moreover, I am part of a department—the General

Secretariat of the Ministry of Justice—responsible for providing digital tools to the entire Ministry of Justice.

At the Ministry, we have a rather broad view of legal professions. There are obviously magistrates, but there are also clerks, prison officers, specialized educators... Justice does not stop at the mere judicial decision. That may be its hallmark, but there is a whole ecosystem where all these professions are interconnected. Our role, within the General Secretariat, is to ensure that cross-cutting issues are well coordinated among the different departments that make up the Ministry of Justice.

As for practical uses, we are not yet equipped with AI tools that are produced, developed, and deployed specifically for use by magistrates in the courts. My comments could therefore end here, but as you can imagine, reality is much more complex and more exciting for tech enthusiasts. First, the Ministry of Justice did not wait for ChatGPT to take an interest in artificial intelligence. We had already launched pilot projects and experiments well before this phenomenon emerged.

These experiments were carried out in the field of predictive justice, also known as jurimetrics, where you take a set of judicial decisions and try to deduce a possible outcome. They did not meet with great success, in the sense that they were not found to be sufficiently interesting in judicial practice. Therefore, they were not integrated into the applications we use. However, when we shift our focus away from France, we see that many countries have embraced this technology and have seen its advantages. Surely, they have also seen its risks and limitations, but for now, let us stay focused on the uses.

Regarding the European ministries of Justice, there are three main categories of possible uses. First, access to the law. Some countries, such as Spain and Portugal, have decided to use a chatbot to provide legal information to citizens. It is a conversational agent to which the citizen will expose their case in natural language. Rather than having a website where you do not know where to click or where you might face difficulties in determining whether it is civil or criminal law, the chatbot will guide you toward the best procedure to follow in natural language. It is not legal advice; it is simply information, but it makes the process more accessible for citizens.

Next, there is the translation function. AI really works wonders in terms of translation, even though one might argue that the meaning and spirit of words could be lost along the way. However, for simple translations like “I want to divorce, how do I proceed?”, it works very well. If you encounter a case where a person is being detained, of foreign nationality, and needs to be informed of basic rules—what will the steps be, at what time meals are served—and they do not speak a word of French, it is useful to have a translation tool. Access to law can be improved using these technologies.

The second category would be that of judicial administration, the way you organize your work within the jurisdiction: sending summons, the orientation of files, transcribing hearings... All these tasks are administrative; they are not decision-making. Various countries, including France through the Court of Cassation, have developed tools for these purposes. For example, the Court of Cassation has deployed an algorithm aimed at anonymizing judicial decisions as part of making these decisions open data. Additionally, it has deployed a tool for the orientation of supplementary briefs—the equivalent of a referral. The AI pre-sorts the summons or brief based on the chamber. There are not many of them, but at least AI does this preliminary sorting, after which there is a human check.

This judicial administration can also be observed, for example, in Germany, in the identification of serial cases. These are mass litigations. If you take each case individually as it comes, it is difficult to group them. Artificial intelligence helps identify them, and process them similarly, either by the same judge or the same chamber.

And finally, there is the famous judicial decision. And here too, things are much more complex than they appear. In the judicial decision, we can see the judge making their decision. One might input several criteria and items, and a machine would then write the legal solution, ultimately substituting the judge with artificial intelligence. In fact, things are much more nuanced. When you are in a decision-making process, it is not monolithic. The solution is not instantaneous when you receive a case.

When you take on a case, there is a whole process that unfolds. You follow the case step by step: reading the case file, identifying the documents, sorting them, assessing the probative value of the documents given to you, and

researching the applicable law. Only after all these steps do you finally start the writing process of the decision. You can incorporate AI into each of these steps, which some countries have started doing. However, today, there is no judge who would say that AI performs the entire task.

Some judges use it for legal research, others to produce case file summaries, or even to identify anomalies in certain pieces of evidence. Indeed, such uses are already seen in some European countries, although we do not yet have enough perspective or feedback to fully assess their reliability. In any case, these are all potential use cases, even if they have not yet been implemented in France. However, we do have many experiments currently under consideration or in the process of being applied.

The use of AI cannot be implemented overnight. A whole set of conditions must be met. However, for example, at the Ministry of Justice, we have decided, on a transversal axe, to work on a transcription tool. This tool proves to be very useful for in-chambers hearings. As a judge holding such hearings, you meet litigants directly in your office, outside of a public courtroom setting. Speaking from experience, particularly when serving as a judge responsible for sentence enforcement or as a juvenile judge, there is not always a court clerk present at those times. You meet people face to face, often delivering news that is not always pleasant.



The use of AI cannot be implemented overnight. A whole set of conditions must be met. However, for example, at the Ministry of Justice, we have decided, on a transversal axe, to work on a transcription tool. This tool proves to be very useful for in-chambers hearings.

At the end of each interview, you must provide the individual with a written transcript for him to sign. It is not always easy to prepare it. You must be at your computer, typing everything they are saying, while also engaging with them and sometimes redirecting their comments. The transcription tool completely changes the situation:

you can focus directly on the interaction with the litigant and be fully engaged in the conversation. For us, this is very promising. AI can serve as a support tool in this type of use case, and the same applies for prison counselors or social workers. It is a tool that could be truly cross-functional for the various justice professions under the ministry's scope.

Lastly, the courts themselves have become increasingly proactive on matters related to artificial intelligence and have decided to launch a few experiments. Specifically, in criminal prosecution within the public prosecutor's office, there is interest in developing tools to help summarize certain documents or large case files, thus assisting with the management of mass litigation. There is clearly an ongoing reflection on these matters. I cannot guarantee that these tools will be deployed soon, such is the nature of experiments: to determine whether it works or not.

What is important to understand is that, at least in France, we are taking a forward-looking approach. We are by no means technophobic. We aim to move with the times when it comes to these developments and uses. That said, we are proceeding with caution because the mission entrusted to us is a sovereign one, an essential and sensitive public duty. And caution means testing, experimenting, before deploying any tool in real conditions involving real litigants.

Sophie SONTAG KOENIG: Thank you for these related examples. Marc CLÉMENT, to maintain this duality, perhaps you could tell us about the administrative view on experimentation?

Marc CLÉMENT: There are a few projects at the Council of State (*Conseil d'État*), but overall, I am not under the impression that we are developing many initiatives, at least not for first-instance courts. For the time being, I do not see any tools being implemented. To be perfectly honest, it seems quite complicated to me.

In the tangible work I do, I do not see much room, even for dressing a summary of the case. There is no reason for me to have a case summary without having personally reviewed the file. As a litigant, if you were in front of a judge who said, "I have not looked at your case, but I got the summary", you would likely question their ability to work at their full capacity. So, I find it difficult to identify areas where these tools could be deployed.

For the administrative courts, what seems central to me is understanding how we will approach the oversight of public administration as it increasingly relies on algorithms. We have all had the experience of using websites where there is no identifiable or reachable person behind them. And that, in my view, is the key issue.

Sophie SONTAG KOENIG: I have not forgotten the academic perspective—we will come back to it shortly. But perhaps, to conclude this segment on the judicial side, we can hear the opinions of the lawyers present here, since law firms are starting to equip themselves. However, not all of them are equally equipped with AI tools, nor do they use them in the same way.

Pascal ALIX: What a broad question. There are tens of thousands of lawyers in France, and hundreds of ways to use tools depending on the profession. There are different types of lawyers, advisory roles, representation roles, and assistance, which are not the same thing at all. Moreover, there are dozens of sociotechnical systems that we can use as different tools, with distinct goals and functionalities. I believe we cannot treat them all the same.

But, unlike the administrative jurisdiction, lawyers are driven by a need for productivity because our law firms are businesses in the end. This might differentiate us from judges. Even though, for public action, there is a certain fear of missing out and not seizing the opportunity, with perhaps a government obligation to implement systems that increase productivity. However, that is not my topic, as I am neither a judge nor a civil servant, so I will return to what I know personally, which is being a lawyer.

For us, it is almost an obvious fact. We must manage a law firm in a competitive sector. If we have competitors who use these sociotechnical systems—which I deliberately avoid calling “tools”—and who know how to use them properly to work faster, generate documents, tables, summaries, translations, and in doing so save time and, in some cases, charge less for the same services, this inevitably creates a form of pressure to use these tools as well.

And this is really the essence of the language, the narrative of the CNB’s (*Conseil national des barreaux*) recommendations: to tell us that adopting these systems is a necessity. It almost



If we have competitors who use these sociotechnical systems—which I deliberately avoid calling “tools”—and who know how to use them properly to work faster, generate documents, tables, summaries, translations, and in doing so save time and, in some cases, charge less for the same services, this inevitably creates a form of pressure to use these tools as well. And this is really the essence of the language, the narrative of the CNB’s (*Conseil national des barreaux*) recommendations: to tell us that adopting these systems is a necessity.

becomes mandatory. I understand that we do not all share the same view on this. My case is somewhat particular, as I am both a professional practitioner and a researcher, and so my critical thinking is immediately alerted. I know what lies behind these systems, so I have a more personal perspective that calls for greater discernment. I encourage reading Yannick’s work, which highlights the practical aspect of using these devices. We need to know what we are talking about, what is hiding behind these interfaces.

We need to understand the implications not just from a professional standpoint, but also on a human level. We are human beings, and therefore susceptible to influence. I was truly struck by a talk given by Murielle POPA-FABRE on latent persuasion—it left a lasting impression on me. This is a message I now try to widely share within the legal profession: these systems are not neutral technical tools. We do not fully understand what we are using—even when it is something like Copilot integrated into Microsoft 365. It is not just a simple word processor. Behind it are algorithms that can influence us.

Why? How? And to serve whose interests? We do not have clear answers to these questions.

We roughly know how language models are trained, we know approximately the techniques, but we do not understand yet how a given result is obtained at any given moment.

Sophie SONTAG KOENIG: Aurélie KLEIN, you are responsible for digital innovation at your firm. You may have concrete examples to address these concerns and points of caution?

Aurélie KLEIN: The use of artificial intelligence—and generative AI in particular—is almost a personal matter. There are those who openly embrace it, those who use it but will not admit it, and those who firmly oppose it. In any case, it is important to reflect, observe, test, build, and deconstruct to better understand and tame these new tools.

At the FIDAL law firm, we have developed and co-designed an internal AI system. And I want to emphasize this idea of co-construction, as we are working hand in hand with a French IT services company, Sopra Steria, to develop these tools.

The idea of developing an internal AI solution within the firm quickly became an obvious choice in our thinking, for several reasons. The first is that we belong to a profession governed by ethical and professional conduct rules. And using off-the-shelf generative AI without paying close attention to confidentiality and data protection is extremely risky, because AI is highly data hungry. We have already discussed the energy and environmental aspects during this conference, but the issue of data is just as critical. AI needs training and input; and it feeds on our data, our training efforts, and our prompts.

The database, which serves as the raw material for generating responses to our prompts, is much smaller than that of off-the-shelf generative AIs, but it is much more controlled. We chose to rely primarily on the BOFIP, a large portion of Légifrance, and a selection of our internal resources. These are documents we carefully selected and validated through a collaborative process, organized by topic and sector of activity. It must be said that in law firms, we tend to be quite individualistic — collaborative work is not exactly second nature. But here, AI has sparked a shared project, one that creates debate and encourages us to discuss how we work and how we use these tools.

To move on to concrete examples, I would distinguish between two main categories of use.

The first includes office-related tasks, in which there is no actual creation of content. It involves formatting, rephrasing, translating... In these cases, we do not have major concerns about content quality, errors, or the so-called hallucinations. AI is simply a form of assistance.

The second category—which is more open to debate—involves content creation and is closer to human reasoning. A lawyer also plays a role as an advisor and strategist; they are not merely a producer of legal rights.

Thus, when defining legal uses and practices, it is essential to rethink and reevaluate the human relationship between the advisor and the client. It is important to provide reassurance, and to demonstrate practical and analytical intelligence. AI can certainly help save time, but its use must remain thoughtful and measured.

Sophie SONTAG KOENIG: With these remarks, you naturally transitioned towards the topic of training. The timing is perfect to address the perspective of professors and students' relation to the uses of AI. So, I am now turning to you, Frédérique BERROD, to share your insights on this subject.

Frédérique BERROD: Following on from what has already been said, generative artificial intelligence raises several challenges for universities. We witnessed it very quickly in Strasbourg, since just one month after the introduction of ChatGPT, we had the first case of exam fraud using this tool. Solutions had to be found rapidly, and this initially generated a degree of anxiety among lecturers. This erodes trust in an assessment that is not carried out under exam conditions, where students are monitored, with no access to their cell phones or computers.

This obviously calls for a whole series of reflections, some of which were conducted at the end of January 2025, as part of a seminar organized by Strasbourg university's training team. Quite unusually, I met between 300 and 350 colleagues there, which just goes to show how much Professors are questioning their practice.

This first element forces us to rethink and reflect on the test itself. What do we ask of a law student to be a good student? At present, what is required is, ideally, to have a complete brain, i.e. to have studied both private and public law, and therefore to be able to think with these series of two major legal disciplines. Generative artificial

intelligence challenges this first presupposition. Interestingly enough, and this was apparent from several of the responses, one of the elements will be to evaluate our students in terms of their ability to do better than artificial intelligence.

In other words, be strategic, critical and probably creative. And I am not sure that creativity is what we are promoting in law schools. It is not what we typically look for when grading a good paper. What is expected of law students is that they should be able to perceive trends in case law, identify the applicable law, and fit facts into categories. And in this respect, they are quite like medical students, who are also confronted with artificial intelligence, as some of my colleagues and I have discussed.

Two problems arise in medicine. The first, which has already been mentioned, is that of the patient who arrives, knowing better than his doctor what he is suffering from, and who claims he is entitled to antibiotics. Until the antibiotic is prescribed, the patient changes doctors. This is a familiar experience for doctors. The same situation could arise with lawyers. As long as the litigant has not found a lawyer who agrees with what he has seen on ChatGPT, I think he will keep looking, even if, at some point, financial means may come to discipline the thing. The other side of the coin that brings lawyers and doctors together is diagnosis. A doctor is trained to recognize a list of symptoms, which he then associates with a list of illnesses. And somewhere along the line, we do the same thing in law: we prequalify the facts a lot.

Which leads me to say that one of the things that should certainly be taught to law school students to develop this creativity, is the ability to listen to one's client. This is probably a problem of organization of the syllabus, because at present, a good law student has taken legal subjects, and sometimes a little bit of management and economics, but not too much, because they need to succeed in, or at least be assessed on, every subject in either private or public law (rather than both).

And these subjects will have to be completed by skills that are not strictly legal. This shift is challenging for Professors. Focusing on students' skills rather than assessing them on their knowledge is a mantra heard a lot in higher education, but not easily applicable in law schools, which remain closely tied to the validation of knowledge, itself necessary for mastering legal



This shift is challenging for Professors. Focusing on students' skills rather than assessing them on their knowledge is a mantra heard a lot in higher education, but not easily applicable in law schools.

reasoning. But, in truth, it is complex to understand what types of skills should be valued.

This issue seems fundamental, since it is becoming imperative that our students are trained in artificial intelligence: they must understand its possible uses, and so on. If I am right about what is happening in law firms, this development is likely to profoundly alter the very nature of the internship currently required to validate a degree. Up to now, the tasks entrusted have mainly consisted of case-law research, actualization work and so on. I fear that AIs, at least in this sector, are on the way to becoming faster and more efficient at these tasks. This raises a series of questions: what is a trainee in this new context? What is their contribution, compared with that of artificial intelligence? And, finally, what professional skills will the student acquire in such an environment?

We also need to think about AI technical training in universities. The government has already given us an answer, in part, by asking that students be made aware of the challenges of sustainable development and ecological transition, as well as the fundamentals of artificial intelligence. Nevertheless, a day is only 24 hours long. Perhaps one day we will find a way to extend this with digital duplicates, but on a more serious note, the real question is how to integrate these new training requirements into already dense curricula.

And, finally, to achieve this, I need to be properly trained in these issues, as a Professor, which does not happen spontaneously. I have taken a personal interest in these issues because I am lucky enough to be surrounded by people who can explain various things to me, such as the construction of a law like the HADOPI law. This raises

questions about our ability to understand digital reality, and we need to be trained too, which represents a real challenge.

Thereafter, everyone will be able to take advantage of these tools and make whatever use of them they wish in their own field of practice: courts, law firms, etc. However, it is vital that the talent pool you draw on is able not only to understand the transformations underway, but above all to keep the human element at the heart of the process. Otherwise, it will quickly become tempting to rely on artificial intelligence to find the solution to a practical case, a judgment commentary or an essay, to mention the classic exercises of legal training—which are probably not the best exercises to prepare for these issues either.

As a result, we are invited to take part in this reflection, but extremely quickly. And perhaps that is what makes it so difficult. I am fascinated by the speed at which these tools are developing, but it also raises a challenge in terms of training. How can we succeed in adjusting our practices, knowing that our accreditations are given for five years? We need to ensure the adaptability of both Professors, as well as students. It is an extremely stimulating and slightly dizzying thought process.

Sophie SONTAG KOENIG: Needless to say, as a lecturer, I obviously share your observations. If only in terms of the changing skills of our apprenticeship and internship students, this is already a very concrete observation. Following this overview of uses—which vary according to profession and subject—we will now turn to the framework within which these practices are taking place: regulatory, deontological, ethical...

I will start by asking lawyers whether European regulations constitute a sufficient and appropriate framework for your practices. In your opinion, are there any internal regulations that would complement this framework? We will start with the general framework, bearing in mind the European exception, namely the regulation on artificial intelligence. However, we should also not overlook the GDPR, which raises many concerns as well, particularly regarding the anonymization of judicial decisions.

Aurélien KLEIN: For our firm, one of the first features developed in our “private” generative AI tool was the anonymization of our documents to facilitate automation and the reuse of certain files.

To regulate the use of AI, there are of course European texts such as the AI Act and the GDPR, but also soft law and best practices. We have discussed the uses and the importance of observing and reflecting before arriving at regulation. I imagine many of you have already become familiar—at least I hope so—with the Artificial Intelligence Act. It is true that some of its provisions remain quite general, with regulation based on risk assessments for certain prohibited uses. But the question remains: what about all the solutions, practices, and usages we deploy daily? Every organization must ask itself whether it is using AI, and if the answer is yes, it must consider implementing a charter of best practices.

There are already proven risks that raise issues of liability. For example, we have seen situations where consultants provided advice entirely based on AI-generated results, overlooking a major cybersecurity issue. From a reputational standpoint, the consequences can be particularly damaging. Just because an AI-generated response is written fluently and without spelling errors does not mean it is necessarily accurate. In such cases, the expert—if they can truly be called an expert—has completely abdicated responsibility to the AI’s output. Our clients are increasingly turning to us to help establish guidelines for the use of generative AI to limit the associated risks.

Starting with this observation, if we consider the context of a law firm, the same requirements apply. We are subject to confidentiality obligations as well as a duty of advice.

First and foremost, it is important to take the time to reflect on these uses, avoid processing personal data, and prioritize human judgment. This principle means that while artificial intelligence can help us gain efficiency and ease tasks, its user must remain the master—not the servant.

One of the risks associated with lawyers or legal professionals using generative AI is the assumption that, thanks to AI, they can now venture into areas of expertise for which they have neither training nor prior experience. Some users believe that AI enables them to explore these “new” fields, aiming to internalize or cover as many legal domains as possible, either to reduce costs or increase revenue.

However, there are many examples showing that there are as many answers as questions for the same legal issue. Reasoning this way is very dangerous. The key point is to understand that



This principle means that while artificial intelligence can help us gain efficiency and ease tasks, its user must remain the master—not the servant. One of the risks associated with lawyers or legal professionals using generative AI is the assumption that, thanks to AI, they can now venture into areas of expertise for which they have neither training nor prior experience.

querying AI in legal professions should only be done in areas and on subjects where we have a genuine capacity for discernment. We must be able to judge whether the response provided is correct or incorrect.

This warning is especially addressed to students, the future legal professionals, but also to some practitioners who may be seduced by this illusion of a magic wand that artificial intelligence might represent.

Sophie SONTAG KOENIG: Earlier, I mentioned the recent Senate report that highlights distinctions between legal professions. Regulated professions benefit from certain guidelines provided by their governing bodies. In my remarks, I was thinking specifically of the framework offered by the French National Bar Council (CNB) on generative AI, via a practical guide published last September. This sets them apart from other professions, which do not necessarily have such guidance. Would you like to respond?

Aurélien KLEIN: Among lawyers, there is a considerable diversity of practices, which explains why some legal organizations are reluctant to define overly restrictive guidelines. What is certain is that we are all governed by ethical principles to which we must adhere, regardless of our practice. Even though anonymization can be time-consuming—if it indeed causes delays—it is mandatory to do it. Laziness has no place here.

Pascal ALIX: I am going to spend a few minutes on regulation, the contribution of soft law and ethical rules, since it is the subject of my thesis. First question: is it possible to regulate this elusive object? I am not even sure after 100 pages of Title 1 of my first draft of my thesis project. Regulating this temporary, ephemeral, elusive object that is generative AI based on LLMs, and which perhaps, if we are to believe Yann Le Cun, will no longer exist in a few years' time, seems complex.

We are building business models, deontological rules, soft law rules, hard law rules and a techno-scientific-legal monument that does not settle everything, the AI regulation—a text that blithely passes over the real cognitive anthropological issues. You can go through this regulation, a veritable gas factory, looking for provisions governing cognitive issues: you will find absolutely nothing. A few transparency measures have been added, belatedly, after the ChatGPT phenomenon, in Article 50, but, in fact, they resolve very little.

That is where we stand on the regulatory front. I hope that those who contributed to writing this text will not hold it against me, but I always hold the same discourse in all my speeches. This text is the legacy of a shift that took place in the 1980s, when we favored *ex-ante* regulation based on a principle of accountability. And speaking as a data protection officer for 12 years, I see the limits of this model daily, particularly when it comes to guaranteeing human rights. In truth, this approach does not protect much at all.

I am also a certification auditor. In this capacity, we rely on a fine reference framework, we check whether such and such a provision is present, whether such and such a document exists. But how can we use it to contextualize and guarantee respect for human rights, so dear to the hearts of many of us in this room? I honestly do not know. And when I say, "I don't know", I think you understand what I mean. Truthfully, I have serious doubts—to put it mildly—as to whether the protection of citizens, administered, of the people who are affected, is really guaranteed. That is the current state of the law.

What we need to understand is that regulation has three main objectives. The first of these is the promotion of artificial intelligence, which can be explained by the geopolitical and economic stakes involved. In fact, this is the primary objective when we examine the preparatory

texts. Having said that, the preamble reiterates the need to respect the Charter of Fundamental Rights more than 150 times.

But if we look at the concrete normative provisions, what do we really find? There are a few elements here and there, including practices prohibited under Article 5. But when you look at the day-to-day implementation of the high-risk systems listed in Annex III, it is hard to conceive how these provisions guarantee genuine protection for citizens, administered and the people concerned. There is very little evidence of this. I do not know if I am overstepping the mark here, but that is the observation I have made.

I simply wanted to remind you of the essential issues at stake. As lawyers, defending human rights is an integral part of our DNA. I thus think it is important to bear this in mind, before we commit ourselves too quickly, and perhaps without sufficient discernment, to the massive use of these algorithmic tools, which carry with them a real risk of dehumanization.

Sophie SONTAG KOENIG: The transition to the question of AI and legal decisions is a natural one, with two viewpoints that can be cross-referenced, thanks to you Claire STRUGALA, Pascal ALIX.

Claire STRUGALA: I would like to make it clear, even though I took part in the negotiations myself, that any decision taken on behalf of France is the result of an interministerial process. It is not, in any case, individual agents taking up the pen to rewrite a text as they please. Negotiating a treaty or a European or international text is an extremely complex process. It is a work of diplomacy and politics. The law comes into play here only from a technical point of view. Of course, it is the lawyers who then translate all this into legal terms, but at the end of the day, it is always about politics and diplomacy. That said, for those interested in international law, it is a fascinating field.

We have been talking about the famous AI regulation. Before even discussing this text, which may—or may not—have reassured some, and which arouses both fascination and concern, I would like to reiterate an essential point. When it comes to judicial decisions, we did not wait for this text to regulate the use of algorithms. Since 2016, algorithms used in the public sphere have already been subject to regulation. This applies

especially to those that might be used by the Ministry of Justice or magistrates.

Whether through the 2016 Law for a Digital Republic (“Loi pour une République numérique”), the GDPR or the French Data Protection Act (“Loi Informatique et Libertés”), the legal framework was already quite clear on the fact that no judicial decision is allowed to be fully automated. In other words, there can be no replacement of the magistrate in judicial decision-making by total algorithmic decision-making. Today, moreover, we are no longer talking about automated judicial decisions, but about decision-support tools. Things are a little more subtle. We shall see that, even if the tool is integrated into the process without entirely replacing it, there are still risks.



There can be no replacement of the magistrate in judicial decision-making by total algorithmic decision-making. [...] Therefore, the AI Act should be viewed as product regulation, not as a measure intended to protect fundamental rights.

All this to say that we have been talking mainly about the AI Act, but it is worth pointing out that the GDPR is also a founding text, just as important as the AI Act when it comes to artificial intelligence. Indeed, without data, there is no artificial intelligence, at least not in the field of justice. Data from your proceedings and court rulings are the raw material for AI in this sector. The GDPR, therefore, is of fundamental importance. Whether it is the AI Act or the GDPR, it is important to understand that there are two sets of regulations: one that applies to justice in the pure sense of the term, i.e. to the judicial decision, and the other that applies to the others.

When I speak of “others”, I am referring to lawyers, in-house counsel, court bailiffs, clerks, and all other regulated legal professions. As far as judicial decision-making is concerned, it is easy to imagine: our terms of use and data processing practices are much more strictly regulated

and monitored than those of the systems and AIs used by what could be described as the private sector of the legal and judicial world.

Thus, the AI Act does indeed constitute a form of regulation, as has already been mentioned. Fundamental rights seem somewhat sidelined in favor of what the text seeks to protect: namely, innovation. As a matter of fact, the AI Act is a product-based regulation. Its primary goal has been to ensure that products placed on the European market—like toys or medicines—comply with certain safety standards and offer traceability in case of malfunctions. Therefore, the AI Act should be viewed as product regulation, not as a measure intended to protect fundamental rights.

AI is clearly now the backbone in economic terms. It represents a key economic, political, and geostrategic issue in which Europe constantly feels it is lagging, behind China and the United States. This is why innovation is a central pillar we seek to protect. And here comes the same old tune: if you want to be innovative, you must avoid imposing regulations. We may agree or disagree with this, but the fact remains: if we want to follow the U.S. model, this is how they have translated the incentive to innovation. Regulation is therefore perceived as a brake on innovation, because it introduces controls, certifications, and prior authorizations before a product can be brought to market. As a result, those who are not subject to these rules will be more competitive.

That said, we are in Europe, and we are committed to preserving our fundamental values. Therefore, we have chosen to reaffirm and protect them by adopting a risk-based approach. Only those systems that raise the greatest risks to our fundamental rights, health, and safety are subject to specific rules and obligations. I am not referring here to prohibited systems, which, by definition, are banned and therefore not regulated. Rather, the systems that are regulated are those that present significant risks—they are allowed but only under strict conditions.

The core of the AI Act lies in setting out technical obligations aimed at securing these systems. However, from the Ministry of Justice's perspective, it is important to clarify that we are not data science experts. That is not our primary area of expertise. We are not the ones developing algorithms, nor are we becoming producers or distributors of artificial intelligence. We could choose to take on that role if we decided

to develop a system ourselves, but that is not our priority. What raises more questions for us are the obligations placed on the “deployers” as defined by the AI Act, that is, the users of these high-risk systems.

The question we can ask is this: does the justice system fall into the category of high-risk systems? The answer is both yes and no. In the context of protecting fundamental rights and identifying what poses the greatest threats to them, we have determined that only systems used to assist judges in decision-making should be considered high-risk. Everything related to lawyers, clerks, and legal advisors does not fall under the high-risk category. As a result, they are not subject to any specific rules, which is why we return to codes of practice for those professions.

In fact, there is no strict regulation regarding the use of artificial intelligence in the legal field, except for the systems used by magistrates. Why is that? Because judicial decisions have direct legal effects and can impact fundamental rights and freedoms. This is the very principle of the force of *res judicata*. That is why we have chosen to ensure that AI systems used in this context are designed in compliance with a set of technical standards.

To put the context of the text into perspective, the AI Act imposes several obligations, both at the development stage and at the stage of use. And who is the user in this case? It is the judge. So, what obligations will we have to comply with if, tomorrow, we use a decision-support tool? The first requirement is comprehensibility. One must know how to use this tool, and I find it simpler to use the word “tool”, even if we do not all share the same terminology. This tool must be understood by the magistrate who uses it. Most importantly, the litigant must be informed about its use.

This brings us back to what was already provided for by the 2016 Law for a Digital Republic (“Loi pour une République numérique”): an obligation of transparency. The magistrate must comply with it, which means, at the very least, indicating whether an artificial intelligence system was used. I have already explained what the Ministry official uses are earlier. The biggest question and the greatest risks—and this is also true for lawyers—do not lie so much in the official use of AI by justice professionals, but rather in its “wild” or unofficial use, done without ever disclosing it.

Today, I cannot guarantee that no magistrate has ever used ChatGPT or similar tools, whether for legal research or drafting purposes. Yet the AI Act introduces a very clear requirement: it imposes an obligation of transparency and understanding regarding the use of such tools.

One final point to raise is the necessity of human oversight. And this is precisely where the interrogations begin to multiply. On paper, this requirement seems relatively straightforward. But in practice, how will a magistrate be able to explain the functioning of the algorithm they used? In what way should they inform the litigant that they relied on an algorithmic tool? Should it be done during the hearing, in the reasoning of the decision, or perhaps through a reference to a website indicating that the Ministry uses algorithms? What exactly does human oversight require?

Because we all know it—despite being told that artificial intelligence is merely a decision-support tool—the law is incredibly complex. When conducting legal research, we are inevitably influenced, to some degree, by the results presented to us. And when we turn to AI, it is usually not to double-check every piece of information it provides; we use it to save time. At some point, a choice will have to be made. It is impossible to verify everything. What exactly are we expecting from this so-called human oversight? What does it involve? And will we truly have the capacity to exercise it?

We also know that humans, in this type of situation, tend to be influenced by what is known as an anchoring bias. And this is even more true for young students like you, young professionals in the making. When you are asked, for the first time, to draft a decision or conclusions, you will not necessarily be confident in your choices. And if you use AI, when you are uncertain, and AI confirms that solution, it is highly likely that you will end up aligning with the machine's opinion. This is entirely logical and normal, it is human.

And when you become a magistrate—if that is your choice—you will find yourself with a pile of cases to process within a very tight deadline. In these conditions, even with the best intentions, you will tend to follow what the machine tells you. And this is precisely the challenge of so-called human oversight. How can we ensure that the tool is useful, helps us save time—because otherwise, what would be the point of using it—while still maintaining control over it?

It is with this in mind that the AI regulation already establishes a set of essential rules. Personally, I believe in the necessity of regulation and in the role of law, sometimes more so than in ethics, in fact. But it also raises many questions. There are many unknowns regarding the implementation of these principles and rules, and about how to concretely ensure these concepts of human oversight, transparency, and explainability.

Aurélie KLEIN: Regarding the legal framework for AI, it is essential to examine the general terms of use of the AI solutions we employ daily.

Pascal ALIX: Just to quickly illustrate what has just been mentioned, Bard's terms and conditions in 2024 summed up the situation very well. They described the tool as recreational and emphasized its exploratory and unfinished nature. This is, I believe, the only time I have seen truly honest terms and conditions that admitted the tool was exploratory, useful for recreational purposes. They warned—and I am not making this up—not to use it for professional purposes because the tool can make mistakes, and therefore, one should refrain from using it to generate results with significant impact.

Sophie SONTAG KOENIG: Marc CLÉMENT, do you share this point of view from an administrative standpoint?

Marc CLÉMENT: To begin with, I believe it is important to reiterate, as I have already mentioned, that currently, I do not see any truly concrete use of artificial intelligence. We quickly imagine a somewhat fantasy scenario where the judge has a tool into which they enter their case, and it then produces a judgment that the judge only needs to sign. But let us be clear, this does not exist. And it has very little chance of existing if we maintain the idea that practicing law is not just about searching for case law.

I completely fathom that, in a training context, this is a significant issue. It is a real question to consider. For example, if I ask Gemini: "What is abuse of rights in tax law?", I will get an answer that, overall, makes sense and is not completely absurd. But if I go further and ask for specific case law references, that is where it completely falls apart, with answers that are completely irrelevant and false. And for me, as a judge in a tax

chamber, the question is not just about knowing what abuse of rights means in theory. What really matters is analyzing the actual business situation, understanding what really happened, financially, with the transactions that took place. All this concrete analysis work, I do not see it being carried out by a machine today. Of course, when we ask general questions, we get general answers, but not precise answers. So we certainly have not reached that stage yet.

I think we really need to go back to this point, because as Antoinette ROUVROY pointed out, the law always steers from an abstraction of the real world. We never directly grasp the world as it is; we transform it into legal categories, into concepts on which we will then reason. So, if we stick to a purely case law-based mechanism, then yes, it is possible to conceive of tools capable of identifying case law. But, if we take concrete examples, such as the one from the FIDAL firm, we see that these are much more targeted and precise tools in their use. They therefore do not raise an issue related to the general use of an intelligent machine, but rather to the specific use that is being made of it.

And this is where the great difficulty of the AI regulation lies. The European Commission built this text on a judicial base of "internal market", with concepts like innovation, products, etc. And this cannot work because we are not dealing with products that we will simply use. The judiciary and law firms are not going to take a tool produced by OpenAI or others and install it directly. We are going to develop applications, and this is not specific to AI. If you look at computing and its developments in the jurisdictions, we do not have the same systems between judicial and administrative systems.

The case management system in the administrative courts is developed by the administrative judiciary itself. The police and the national gendarmerie do not use the same systems. They have tried, through considerable spending, to harmonize things, but without success. We cannot imagine an AI tool, developed for police work, appearing on the European market—say, made by Estonians—and the French police saying, "This is perfect, it is exactly what we need, let us take it and implement it". We are not operating through product-based logic, but rather through use-based logic.

And it is in the context of these uses that we must begin to define ethical questions, in the

sense of how decisions will be produced. And this ties into the central issue of decision support. And once again, I do not see it so much in the context of judges, but rather in administrative decision-making. What does it mean for a human being to have been involved in the process of producing that decision? Habits form over time, and gradually, we either trust or do not trust this process.

We are faced with a significant level of complexity, namely having to determine whether human intervention is necessary. And if it is, what exactly must humans do, and how can we ensure that they do it properly? Suppose we establish an administrative law rule requiring that the person who makes and signs the decision must have reviewed the file. How can we guarantee that this was truly done? The ethical rule will not concern the quality of the decision, the outcome, or whether the decision is fair or legal. It will simply be about ensuring that the administrator, the civil servant, has really done their job. This is one of the key challenges and questions that need to be addressed.

Sophie SONTAG KOENIG: In addition to what has already been said, even if we are stepping outside the strictly judicial field or the professional practice of law, at the university level, we face similar reflections regarding regulation. We know of challenges related to training, which have already been raised, and with mentoring, which we are already aware of within the academic and research communities.

Frédérique BERROD: Perhaps to restate—and this is likely my background in European law speaking—I believe we tend to interpret the AI regulation as saying more than it does, and in areas that are not strictly within its scope. This regulation is not intended to regulate artificial intelligence as such, nor even the products that directly involve AI applications. Its primary objective is risk prevention. It is a regulation that was conceived in a post-COVID context. The idea is first and foremost to prevent risks in terms of health and safety, in relation to Human rights. That is why we constructed the well-known risk pyramid found in the regulation.

Why do I say this? Because it raises an important issue for researchers, which is excluded from the scope of the regulation. This means that within the framework of research, it is possible

to use and produce any kind of AI without being subject to any specific rules. And this exclusion makes sense, since proper research should not be constrained by overly strict regulations that could impact the substance of that research. Nevertheless, this creates a problem, because to uphold European values, we ask researchers to conduct ethical research, grounded in ethical principles. One key issue, then, is figuring out how we will put that into place.



[...] proper research should not be constrained by overly strict regulations that could impact the substance of that research.

And we need to do better with the AI regulation than we did with the GDPR. Today, the GDPR has led to several extremely negative consequences for social science disciplines, because it is being applied without fully leveraging its potential. Everything is seen as contrary to the GDPR before even questioning the scope and definitions. I fear the same thing may happen with the AI regulation, especially given the blurry line between what constitutes research and what constitutes research and development.

As for research and development, once there is commercialization of what has been produced through AI-based research, or research that will result in AI, then the risk pyramid must be applied. In terms of the researcher's responsibility, or even that of the university, it becomes necessary to ask whether a system has been developed that, for example, falls under the category of prohibited AI. This raises a real issue in terms of how such research is, or is not, overseen, and in terms of the viability of the economic model necessary for its development.

We are witnessing a dual line of reflection in education. First, there is reflection at the university level, likely due to what is often—though inaccurately—referred to as university autonomy. Each university is currently considering its own regulations, whether it concerns testing, which I mentioned earlier, the revision of certain courses, or the use of AI. One of the questions that will need to be addressed, and which I find extremely challenging, is that we must teach

with an emphasis on critical thinking. It may seem obvious, but personally, I do not have all the answers to say what skills need to be developed to develop critical thinking in law, at least when it comes to skills beyond strictly legal ones.

Another debate is whether we can trust public data to conduct legal research. Are we going to need to secure certain elements, as we do, for example, with health data? Moreover, we know that today, research on health data is generally not of very high quality—particularly in France—precisely because the research is not sufficiently regulated or does not sufficiently comply with existing regulatory frameworks. This presents a fundamental issue in terms of research content and the geopolitics of research, especially when it comes to health.

We are therefore witnessing the development of AI code of ethics, for example, outside of France, such as at the one from the University of Geneva. Additionally, there is a more collective reflection at the European level, carried out within the frameworks of university collaboration. I am particularly thinking of the League of European Research Universities, an association that brings together universities to jointly reflect on European issues that go beyond strictly French concerns and that will certainly need to be considered.

The impression I have, from listening to my own needs and those of my colleagues, is that we need good practices to help us navigate better. Otherwise, the tendency is to reject AI because we only perceive its potential dangers. Which, in my opinion, would be a mistake, but it is one of the risks nowadays. This code of best practices, which you mentioned in the context of regulated professions, could be a good idea for universities as well, just as feedback from experience could efficiently guide our interrogations.



This code of best practices, which you mentioned in the context of regulated professions, could be a good idea for universities as well, just as feedback from experience could efficiently guide our interrogations.

I believe that, in the face of ChatGPT, we have moved from a period of astonishment—which led us to prohibition—to a period of excitement about the use and teaching of AI, with a strong appetite for experimentation and experience. This indeed stems from the significant issues I tried to outline earlier, and which continue to raise questions among many colleagues. At least, that is how I am feeling at this stage.

Sophie SONTAG KOENIG: We are moving towards the last moment of this round table. Several points have been addressed that touch upon the question of misunderstanding. This was the central theme we had agreed to discuss, which you spontaneously tackled. It was originally planned to be covered towards the end, but you have incorporated it quickly, as the “misunderstanding” is palpable. You have mentioned it in terms of disappointment regarding the initial expectations of artificial intelligence, but also from the perspective of tangible risks.

Other points have not yet been addressed, and we will now be able to focus on them for these concluding remarks. We promised ourselves to cover three questions, including the issue of proof, in its use by judges. What will happen to proof if we use AI as evidence for judicial decisions, whether to strengthen its probative value or to question it?

Claire STRUGALA: Since we are talking about misunderstandings, I think that when it comes to AI, it is very difficult to form a clear opinion because there are numerous technologies, many different uses, as well as various concerns and expectations. It is a fascinating topic, but difficult to grasp, and even more so in a round table. I will therefore briefly address the question of proof, as it seems to be a topic that is not often discussed in relation to AI and justice. Then, I will mention two misunderstandings that I identify at my level.

Regarding the question of proof, when we talk about AI and justice, we always come back to judicial decision-making. The risk is that AI might capture the decision-making will of the judge, and we focus a lot on that. However, there is a whole issue that is not often discussed. Today, and you are all aware of this, among what you consult on the Internet and receive on your phones, we hear a lot about deepfakes. We know for a fact that all the data we receive may have been partially or entirely generated by AI. We

know that data can have been augmented by artificial intelligence. For example, low-quality surveillance camera footage can be enhanced using AI technologies.

We know that in healthcare or in many other disciplines, judicial experts use artificial intelligence in their professional practices and in their expertise. Therefore, in the short and medium term, we will see procedural documents that have been provided to us by the parties themselves to support their claims, and we will no longer know their probative value or, at least, their origin, which could lead to major difficulties.

If we are not informed, we could be deceived by technology that has been used. For example, someone could provide us with a fake audio recording of their ex-wife or ex-husband in the context of a divorce procedure, or someone could submit a fake image of their neighbor dancing naked in their garden, claiming it is an invasion of their privacy. These are situations that will arise, and we are not at all prepared for them.

Lawyers are not prepared for this, because they will need to ask the client about the origin of the evidence they provide. Experts are not prepared either, because we have not told them that it might be necessary, in their expertise, to specify that they used AI and potentially mention which AI they used. The adversarial process should allow us to challenge this AI. Finally, judges are not prepared either, because this has never occurred before, and they will have to decide for themselves what probative value to assign to this piece of evidence included in the case. And if we, as judges, are not yet ready to take the big leap, litigants and citizens already are. This is therefore a major issue, as we must assess our ability to receive these new pieces of evidence.

Regarding codes of practice, we should collectively develop guidelines based on a simple yet time-tested principle: transparency. If you talk



Regarding codes of practice, we should collectively develop guidelines based on a simple yet time-tested principle: transparency.

about it, if we are aware of it, it is already the best way to question and, if necessary, challenge the information provided to you.

Regarding the two other major misunderstandings I identify, the first concerns especially the Ministry of Justice, even though I do not speak on its behalf. What is called technological solutionism, is this idea that if you integrate AI into a struggling organization that lacks resources, it will solve all the problems. Today, AI is presented as a productivity tool. It is nothing more than that. It is not about AI being a human miracle or changing people's lives or making them much more protected and happier. AI helps to save time, that remains the main goal.

The report of the General States of Justice, the Senate report recently released on AI, and the solemn opening speech by the first prosecutor and the first president of the Court of Cassation all confirm this: the justice system is struggling, and it lacks resources. Consequently, AI could be a solution to improve the work of the clerks, or to make magistrates' work more efficient. However, in my view, this is a technological and solutionist myth, although I will not anticipate what Pascal ALIX will certainly discuss regarding biases, difficulties, and risks.

We are aware that AI is not a miraculous tool. As an administration, we must first realize that the Ministry of Justice needs to catch up on its digital lag. AI is a sports car, a Ferrari. But if you start with a slow car, such as the Renault 2CV, there is still a very long way to go. We really need substantial resources to invest in our courts to be able to welcome these technologies, and we are not quite yet ready, because all of this requires a lot of resources.

Moreover, to implement AI, one must have access to data and ensure the security of Internet hosting. As a sovereign ministry, we are subject to cybersecurity requirements, and we must offer secure Internet hosting in compliance with the very strict rules of the GDPR.

Next, professionals must be trained and supported. It is easy to advise "to do AI", but if you do not know how to ask the right questions and you do not understand how to draw conclusions from the information provided, you will not gain anything from it. Finally, it is necessary to have budgetary resources and a certain proportionality between the goals you wish to achieve and the resources you are willing to allocate. All of this means that AI can be an opportunity, but it



It is easy to advise "to do AI", but if you do not know how to ask the right questions and you do not understand how to draw conclusions from the information provided, you will not gain anything from it.

cannot be a solution to all our resource problems, whether for the Ministry of Justice or any other administration, for that matter.

The last myth that I wish to debunk is the one that says that if AI is used by judges, we will end up with inhuman and standardized justice. Personally, I do not share this view, as long as, of course, we do not purely and simply replace a human judicial decision with a fully automated one. To begin with, AI is not necessarily inhumane in its interactions. This was demonstrated earlier in a rather humorous way with the student writing her thesis and being encouraged by the AI, which said: "Go ahead, you can do it, do not give up". AI, depending on how you interact with it, can be more compassionate than your professor or your own relatives. This is the first point.

The second point is that, if tomorrow, AI is used as the basis for decision-making, meaning it will propose a median solution, I would say it will ultimately push us humans to improve. How? In a legal dispute, there are always two lawyers: one who agrees with what AI suggests and one who disagrees. The key point here is that the lawyer who disagrees will have to prove to the judge that the solution proposed is not the right one for their client. Therefore, they will have to demonstrate more creativity and delve into their case to explain why the suggested solution is unsuitable. And this will force the judge to better justify its decision and explain the reasons behind its choice.

In the end, claiming that AI will standardize everything depends on how we approach it. If we accept it as a given, perhaps. But if we see it as a challenge that allows us to discuss and better justify our decisions, it can be beneficial.

Furthermore, if I manage to save time administering my cases, it allows me to dedicate more time to listening to the litigant or the lawyer. It is not behind a desk, doing research, that the most powerful and resolute solutions to issues arise, it is often in the courtroom. I thus question the idea that AI will necessarily dehumanize our justice system. We must adapt and ensure that it becomes real support, enabling us to focus better on our core profession.

Sophie SONTAG KOENIG: We could almost say the floor is given to the defense to finish. Regarding these applied myths, we have two representatives from the lawyer profession to recenter the ideas. You have expressed several points concerning the loss of specialization from the lawyer's perspective, as well as the added value of the lawyer, which Aurélie KLEIN has highlighted well.

Pascal ALIX: I will take on the role of a techno-critic, even if that might seem overly simplistic, because I am a heavy user of all these tools, even before ChatGPT, GPT-2, GPT-3, etc. To avoid this first misunderstanding, I want to clarify that I am not just a techno-critic, I am also a user. In fact, I acknowledge that these tools can be extremely useful for creating tables, briefs, and even in the ideation phase, which we have not talked much about, namely, the generation of ideas. But we must keep in mind that this is only one part of the process.

It is not just about pressing a button to get conclusions, a contract, or a judgment. It is the same for law firms. By pressing a button, we may get a snippet of an idea, a paragraph, that we might eventually reuse. I think we have not emphasized enough how these devices work concretely.

We have various technologies that draw on databases, what we call RAG³, which produce fairly reliable results, or at least more reliable than simple text generation as described earlier this morning. There can be a part of text generation and a part of retrieval, meaning the search for sources. That is what a tool like Perplexity does, or tools from publishers, such as Gen-AI-L, mentioned earlier and made available to several law firms.

³. Retrieval Augmented Generation.

Hence, we are confronted with these technologies. Practically speaking, when we press a button, we do not get much. We get a snippet of an answer that may be coherent if we ask a question like: "Describe what such a notion is". It is a general, coherent answer, but we do not use that directly in legal submissions. You need to recontextualize, reason, qualify, and interpret the facts. This is the entire hermeneutic part, which has been barely mentioned so far. AI is incapable of carrying out this hermeneutic work, whether for lawyers or judges. It cannot humanly immerse itself in a case file and imagine, with true empathy, what it is like to be in the shoes of the client, the opposing party, or any of the parties involved.



AI is incapable of carrying out this hermeneutic work, whether for lawyers or judges. It cannot humanly immerse itself in a case file and imagine, with true empathy, what it is like to be in the shoes of the client, the opposing party, or any of the parties involved.

We are aware that these devices, even with the latest developments from Anthropic and the chain of thought technique that produces pseudo-reasoning, are no longer just a stochastic parrot, to borrow a phrase from a previous speech. We agree that there have been developments, and sometimes extremely rapid ones. Since early 2023, things have evolved; I have even been able to submit complex cases to these tools, including Claude 3.5. I am quite fascinated by the results and the speed of these changes, which have made these technical devices increasingly powerful. But there remains this major misunderstanding: the belief that it works like a human being, even though it is merely a simulation of a human with real reasoning.

We obtain a result. But there are two ways to get there. We arrive at something we can reuse after a pseudo "conversation". After all, most of these systems are called conversational agents.

So, we reach a result through iterations and cross-checking. To get to a roughly satisfactory draft, I combine ChatGPT Pro with a customized version, what we call GPTs.

I feed it with numerous internal documents, obviously anonymized (I will not go into the detail of best practices). Anonymization, documentation, and post-checking should normally be clear to everyone nowadays. We can reach something, but it is only a draft. We really need to stop fantasizing about the idea that it will generate a final ruling, a legal submission or a contract. This fantasy persists, even though it only produces a draft that we then rewrite after doing additional research.

I thus come to a second misunderstanding. Since this draft must be rewritten, how much time does it take? Is it really a gain in productivity? I am not so sure. Personally, I have spent about 2 to 3 hours each day, since November 2022, testing different generative algorithmic systems. I am not convinced there is a productivity gain when it takes 50 iterations to finally arrive at something satisfactory, that you then must cross-check with another AI system to verify it, knowing that there are very efficient verification prompts.

You can create verification prompts for the output produced by another algorithmic system to reach a quite impressive result, but of course, in the end, you still are obliged to verify it. So, it is preferable to work with result engines, using RAG, that search for sources rather than generate text probabilistically. On this topic, there are hundreds of articles on hallucinations; it is well known now, we all know it. But even with RAG and databases, we still end up with a result which, due to the generative part, we cannot be 100% certain it is reliable and relevant.

Does this concretely mean that all this will allow us to gain productivity? We are not even sure about that. The problem lies in our capacity for discernment. I can do it now, but only after thousands of hours of testing. How can someone who has not experimented like I have—over so many hours, across different algorithmic systems that evolve “faster than light”—possibly has a capacity for discernment?

I had initially planned to create an empirical comparative evaluation grid, as part of my thesis, to assess which systems are efficient, which are not, and to evaluate reliability, the possibility of eliminating hallucinations, etc. I quickly gave up, because everything changes constantly. I have

read many research papers, and there are indeed reliability comparisons, but they are disputed by other studies. It is endless. In the end, we are in the fog, in the intangible. And I remain in the intangible, even after testing all these models. Yet, another myth.

One final myth, the myth of neutrality, which was already discussed during the previous round table. AI is obviously a sociotechnical object that carries political values, which has been thoroughly explained. However, there is something we have not talked about much that is extremely interesting when it comes to using these tools in research: the issue of non-replicability. Unless I am mistaken, this is one of the conditions for scientific integrity. For a tool to truly support us in our research, it needs to yield consistent results if reused the next day.



For a tool to truly support us in our research, it needs to yield consistent results if reused the next day.

I had asked a question about the EU IA Act: “How many articles does it contain?”. One answer said 95, another said 84, as it was based on—even though the adopted text had been published—the older versions of the text still under negotiation. And as we know, hallucinations depend in part on the training date of the model. But here, even when asking the same question to the same model at the same time, the answers differ. So how can we conduct serious scientific research with a sociotechnical, political object that cannot guarantee replicability?

Sophie SONTAG KOENIG: Frédérique BERROD, as a Professor, and from this academic perspective, we will give you the final words addressed to the students who represent tomorrow’s professionals.

Frédérique BERROD: I would simply like to offer two remarks. It might be worth reflecting on what we could lose by using generative AI too easily. For example, in our daily practice, we are often

asked to condense or summarize texts. Being a good legal professional normally requires a capacity for conciseness. If we systematically rely on AI to do this—which, admittedly, can perform quite impressively—are we, in doing so, losing a skill?

In my view, this is a consideration that still has not been properly addressed. In the pharmaceutical sector, we transitioned to chemical medications, for example, which led to a loss of part of our knowledge on plant-based remedies for instance, and today we are seeing that this is causing problems for some pathologies. Let us avoid making the same mistake when it comes to skills. Listening to you throughout this conference, I would say that the incredible opportunity students have with artificial intelligence is that it likely gives them—more than students of my generation—a duty to go back to the human.

Going back to the human essentially means being capable of three things. First, being creative and not fearing that creativity, even if it does not necessarily conform to a copy-pasted “ideal” answer. Second, it is about curiosity, looking at the world from a 45-degree angle. When you look at the world from that angle, you do not see it in quite the same way. This represents, in

my opinion, a form of intellectual curiosity that is extremely important. And finally, third—and perhaps the most defining characteristic of being human—is the ability to listen. Listening both to society and to your clients, to all those components that help us preserve what we still call, although it is just as mysterious as artificial intelligence, the added value of humans compared to AI. Ultimately, I find this rather exciting.

Claire STRUGALA: To align with this vision, the message I would like to convey is that we must show a certain boldness in using these tools. Without boldness, life would be quite dull. But we must also be cautious about the trust we place in these tools. I believe that boldness and caution are the two key principles expected from a judge at a certain moment, and I will add a third: transparency. Regardless of the professional context in which you will use these tools, be transparent, talk about it with your colleagues, your superior, and with those involved in the case. Do not make reckless uses of AI.

Sophie SONTAG KOENIG: Everything has been said: transparency, caution, boldness—this will be the starting point optimism. ■

Closing conference:

AI, a weapon of mass destruction?

- Artificial intelligence is a marketing fantasy, its origin is not really aimed at reproducing human intelligence but at simulating it
- The historical use of artificial intelligence from the Second World War to its current deployment in autonomous weapons systems is military
- The social and political consequences of automation are to be apprehended from the angle of massive job destruction and the instability of political regimes



Charles THIBOUT

Doctor of political science, member of the "Centre européen de sociologie et de science politique" (CNRS, EHESS, Paris 1 Panthéon-Sorbonne)

Yannick MENECEUR: I would like to welcome Charles THIBOUT, Doctor of Political Science, a member of the European Centre for Sociology and Political Science. I have entrusted Charles with a delicate mission, that of addressing another question—one of my whims—that of "artificial intelligence, a weapon of mass destruction?". You will naturally recognize the reference to Cathy O'Neil and her book "Weapons of Mass Destruction", which highlights the major impact of algorithmic systems on our societies, to the point of becoming weapons of mass destruction. The resulting geopolitical context is complex.

Charles THIBOUT: Thank you all for still being here after such a long day. Indeed, the theme is "artificial intelligence is a weapon of mass destruction". To be completely honest with you, I am obviously not the originator of this title, but I found it thought-provoking, which led me to

keep it as is. To conclude this day full of presentations, it may not be such a bad idea to introduce a touch of lightness, which is reflected not on the substance of the message—the weapon of mass destruction—but more in the sensational form of the title, somewhat like all those magazine covers that have flourished in recent years in our newsstands, promising, depending on the editorial mood, either salvation or the apocalypse. Sometimes an apothecic destiny for our civilizations, and sometimes, more often, the catastrophe of a world where machines will gradually take over.

As the title of this closing conference suggests, artificial intelligence indeed holds significant evocative power, which invites into the discussion a few demons and other ghosts, or *fantasmas*, as the Greeks would call it. As soon as we attempt to define this topic, we find ourselves, unwittingly, conducting a collective anamnesis of an ancient time, where myth and reason, where *muthos* and *logos* were synonymous, and where the supernatural was fantasized as an unknowable reality. It may not be useless to distinguish between fantasy and reality, while taking these sensationalist expressions seriously.

First, I would like to remind everyone of what is probably already well known: the expression "artificial intelligence" itself is the product of a

fantasy designed for marketing purposes. It was with the intention of selling a product and securing funding that this most seductive semiotic packaging was created by the American mathematician and cybernetician John McCarthy, to designate the new field of research he claimed to have founded in 1956, along with others, at the Dartmouth Conference. In fact, he succeeded in his goal, as his scientific and technical intentions had a profound impact on the imagination of his contemporaries. Clearly, the researchers and engineers behind artificial intelligence never truly believed, unlike their cybernetics ancestors, that it would be possible to replicate human intelligence in a machine. Their aim was always simply to simulate it. From the very beginning, the term "artificial intelligence" created a gap, a hiatus between the actual technical project and the much broader possibilities it implies or suggests.

He thus paved the way for truly fantastic hopes, in other words, for fantasies in the psychoanalytic, Freudian, sense of the term. And if this semantic magic trick worked so well, it is because it activates the reminiscences of a considerable mythological chain of representations that trace back to the Greek and biblical foundations of Western societies. These representations evoke the demiurgic act of the god(s), which humanity would strive to replicate to overcome its own incompleteness, to abolish the distance between the creature and its creator, and ultimately to reunite the creating and the created. It is this very fantasy that lies behind artificial intelligence: elevating man to the position of God, killing the father, and appropriating his attributes.

It is the end of the adamic fall, a return to the Garden of Eden for the monotheists or to the Golden Age for the Greeks. Except that there is a problem, which is that, as in the Oedipal schema, the parricidal fantasy has a double effect: it elevates the son to the level of the father, and makes this son, now a father himself, the prey of a new son who will seek to kill him and replace him. Hence also this deeply rooted fear in Western culture, the fear of the creature escaping from its creator, from the monster of Frankenstein to *iRobot*, through *2001, A Space Odyssey*, the machines of *The Matrix*, *Terminator*, and so on. The centuries-long accumulation of these stories has shaped our cultural references and, with them, our chains of perception and thought, which, by default, categorize the artificial creature among threats.

The international campaign against killer robots launched in 2013 by Human Rights Watch, as well as the last statements of Stephen Hawking before his death, who feared that artificial intelligence could end humanity, are merely the latest manifestations of the fear that golems and their modern avatars elicit. Not without ambiguity, since at the same time, these creatures fascinate, both for their exceptional qualities and for the gratifying, narcissistic image they reflect onto human beings, their creators.

This ambiguity is well perceived in Putin's remarks when he declared in 2017 that whoever becomes the leader in artificial intelligence will become the master of the world. For artificial intelligence indeed embodies the fantasy of the instrument of supremacy, the ultimate weapon. But, at the same time, this transforms it into a major political issue. Indeed, no one wants to take the risk of seeing a competing country seize this technology first, since the presumption is that whoever manages to master, passed a certain threshold, the highest level of sophistication in artificial intelligence, will forever outpace and subjugate all rivals.

We therefore observe that all these sensationalist headlines we display with artificial intelligence, to attract attention here today, fundamentally express an eschatological fantasy. This is something we have not seen at this level of intensity since World War II, perhaps, and the race between the Allies and the Germans to be the first to obtain the atomic bomb. When we talk about a weapon of mass destruction, we are indeed reviving something that belongs to the traditional fantasy of artificial intelligence, a fantasy of the end of time, of the end of history. But this fantasy has real implications. And since the title invites us to do so, let us examine it from a military perspective. When did the first attempts to apply artificial intelligence for military purposes take place?

During World War II. Of course, artificial intelligence does not exist as an autonomous field of research at that time, but its precursors, whether cyberneticians or information theorists like Claude Shannon, who was one of the participants at the Dartmouth Conference in 1956, were already producing scientific work related to it. It was two cyberneticians, Walter Pitts and Warren McCulloch, who in 1943 conceptualized what we today call artificial neural networks, which are at the heart of today's connectionist approach to

artificial intelligence, as opposed to the somewhat outdated symbolic approach.

But it does not stop there, as a large part of those who participated in the foundation of the artificial intelligence research field, or who laid the scientific groundwork for it, worked during World War II in the U.S. Department of War, some for the Manhattan Project. Others, such as Norbert Wiener, who was the leading figure of the cybernetics movement, or Julian Bigelow, an IBM engineer and participant in the inaugural artificial intelligence conference at Dartmouth, worked on a project to automate American air defense.

Their project, which we would almost call artificial intelligence today, aimed to predict the trajectory of enemy planes using statistical and probabilistic processing, incorporating various input data such as altitude, speed, acoustic data related to the planes, and adding data from previous observations to refine the model and prediction. Conceptually, this is very close to what we today call machine learning. The military, indeed, is the sector of society that has the most faith in the miraculous possibilities of artificial intelligence. From the early 1960s, millions of dollars were spent every year by the Pentagon's research and development agency to fund the artificial intelligence laboratories at MIT, Stanford, Carnegie Mellon, and IBM.

In fact, these laboratories were created by participants from that famous Dartmouth Conference in 1956. The artificial intelligence laboratories at MIT and Stanford were founded by John McCarthy and Marvin Minsky. IBM's laboratory included members such as Julian Bigelow, already mentioned, Nathaniel Rochester, who was the chief architect of IBM's first commercially marketed computer, which had the NSA and the U.S. Navy as its first users. Since then, artificial intelligence has continually been integrated into the Pentagon's military program. As early as the Vietnam War, Robert McNamara, Secretary of Defense under Kennedy and Johnson, set up a team of computer scientists to apply artificial intelligence technologies to the act of war.

One such computer was developed within this framework, with the aim of providing operational advice to Pentagon strategists based on certain input data, such as demographic, military, and economic information about the situation in Vietnam. Of course, at that time, there was not the technological maturity we have today. Data

was scarce, and computing power was far inferior to what we have now. To give you an idea, when Pentagon engineers asked the computer in 1969 when they would win the war, the computer responded: "You won in 1964". This clearly illustrates just how far the fantasy of a machine capable of mimicking or even surpassing human intellectual abilities can lead.

This fantasy of a scientific government is deeply rooted in the very origins of the cybernetics movement. Cybernetics etymologically means the science of government or the art of governing. Today, artificial intelligence is mainly used in the military domain as a decision-making aid, but it also provides weapons, what are known in military jargon as Lethal Autonomous Weapon Systems. Today, the world's major militaries are developing and deploying this type of weapon. For example, in 2021, the Israeli military used swarms of autonomous drones to carry out military operations in the Gaza Strip. This has been the case again since October 7.

We learned yesterday through the *Washington Post* that Google is providing its artificial intelligence tools to the Israeli Defense Forces to conduct operations in Gaza, which could constitute potential complicity in crimes against humanity, or even genocide for companies that participate. I will close that parenthesis. In South Korea, since 2014, there has been a program to systematically replace human sentinels along the inter-Korean border with robotic sentinels, designed and manufactured by Samsung. Yes, artificial intelligence is indeed a weapon, it is in circulation, and that raises a lot of problems. As legal professionals, you know very well that the use of these killer robots blatantly violates at least the principles of proportionality, discrimination, and accountability under international humanitarian law. But this is already a topic of well-known discussions.

I would rather address the reasons that make artificial intelligence a weapon of mass destruction by design, rather than merely metaphorically. To do so, it is essential to examine what seems to be the main point of articulation in the development of artificial intelligence since its origins, that is, since World War II. This point of coalescence, this unifying motif for all artificial intelligence actors, concerns all its application perspectives, setting aside fundamental research. It has always been the search for increasing returns, for increasing productivity, whatever the application domain.

How can we achieve increased productivity? By aiming for the generalized substitution of machines for human beings. We can turn the problem around in every direction, but this is the core: making production, of whatever kind—industrial, commercial, financial, and even intellectual, social, political—as efficient as possible, by assuming that all other things being equal, it is the human factor that limits the possibilities of unlimited productive surplus. Humans are, indeed, finite, limited, and limiting. If we want to push the limits of performance, achieve the optimum of intellectual capacities, establish the best government ever conceived, a government based on science and mathematical rationality, it is obvious that we must do without humans.



If we want to push the limits of performance, achieve the optimum of intellectual capacities, establish the best government ever conceived, a government based on science and mathematical rationality, it is obvious that we must do without humans.

This way of reasoning, which implies that each human being inserted into the production process represents an unbearable marginal cost, fits quite well within our capitalist regimes. With artificial intelligence, the possible invalidation of the iron law of wages becomes an attainable horizon. It opens the possibility of increasing productivity beyond human capacities, all while dispensing with labor. Labor, of course, is costly, unreliable, and must be fed, cared for, and have its work capacity restored by resting, etc. We can easily imagine the appeal of these types of promises to business leaders, management committees, boards of directors, and shareholders of more or less large companies. The goal is clear: the complete substitution of capital for labor.

This is exactly how the producers of artificial intelligence tools, notably the major companies in Silicon Valley, sell their solutions to businesses

around the world, but also to countries. And for good reason, we now know very well, and it has been thoroughly documented, how the managerial mindset of capitalism has long since infiltrated governments. We think of austerity policies, which have become the norm for budgetary policies, or the rationalist aspirations of new public management, reform policies, and administrative modernization policies that have taken over the bureaucratic space since the 1980s. All of this has created fertile ground for attempts to import artificial intelligence technologies into political processes and administrative functioning circuits.

One need only to look at what is happening in the United States. Elon Musk is about to be promoted to the head of a brand-new department focused on government efficiency. What is his agenda? To use artificial intelligence to identify administrative areas that can be subjected, and I am quoting him, to massive staff and budget cuts. It is also about using artificial intelligence to sift through millions of pages of regulation to identify those that are, again I am directly quoting him, without explicit legislative foundation. We are witnessing a vast regulatory purge in areas such as taxation, energy, environment, finance, etc.

We tend, especially at this moment, to focus on Musk and Trump, but similar alliances between states, government, and American tech also exist in France. My thesis focuses on the relationships between alliances and confrontations between Google and the French government. Here is one example among others: since 2019, the Directorate General of Public Finances (DGFIP) has been using Google's artificial intelligence technologies, specifically computer vision technologies, to automate its tax auditing operations. In plain terms, Google, which by the way entered into a public interest judicial agreement with the tax office the same year, is being used by the DGFIP to detect undeclared buildings—such as pools or facilities—of private individuals, constituents, using images from the National Institute of Geographic and Forest Information (IGN), when not directly using Google Earth or Google Maps.

The goal of this program, called “innovative land management” (“Foncier Innovant”) is to achieve 12 million euros in savings for the administration, and to target, in the medium term, 50% of individual tax audits using artificial intelligence, while eliminating 300 full-time equivalent positions. The late Claude Allègre dreamed of it, Google made it happen, “slimming down

the mammoth". Here, this does not concern the national Education system, but after all, why could artificial intelligence not do better than a teacher? Especially if they are always absent, as a minister recently declared.

We are no longer in the realm of sensational newspaper headlines. Not only do the promises of efficiency, effectiveness, and performance of artificial intelligence perfectly align with the budgetary objectives of the states and the profitability concerns of businesses, but most importantly, this is directly translating, in a very straightforward manner, into job cuts. And we are probably just at the beginning. According to forecasts from the World Bank in a 2019 report, the automation and robotization of work could lead to the net disappearance of 47% of jobs in the United States, 66% of jobs in India, and up to 77% of jobs in China by 2050.

In another, more recent report published last year by the IMF, it is stated that 60% of jobs in advanced economies, that is, Western economies on average, are threatened by artificial intelligence, including 60% of jobs in the United States and 70% in the United Kingdom. It is important to clarify that we are talking about jobs at risk of being eliminated, not replaced or replaceable. In other words, two of the largest international organizations, the World Bank and the IMF, predict that most populations will eventually find themselves unemployed.

Question: if these predictions turn out to be accurate, what will happen to these populations? Will countries meet their basic needs? The most likely conclusion of such a social catastrophe will be the marginalization of these populations. It is understood that states will almost certainly be unable to gather the resources required to guarantee a decent standard of living outside of any form of paid employment. As a result, not only the survival of these populations, but also, quite evidently, the stability and survival of the political regimes themselves are at risk.

We can also consider a forward-looking exercise. If mass unemployment indeed becomes a widespread phenomenon, only those countries whose economic fabric generates sufficient productivity gains to continue ensuring a satisfactory standard of living for their population will be able to escape the resulting socio-political upheaval. That is, of course, assuming that adequate redistribution mechanisms have also been put in place. We then begin to understand what Putin

meant when he said that artificial intelligence would be the key to hegemony in international relations.

If artificial intelligence does indeed fulfill all its promises in terms of economic efficiency and productive performance, and if one country reaches a stage where its competitiveness becomes unmatched, it will be the only one capable of bearing the economic, social, and political costs of mass labor displacement. Other countries, lacking this competitive advantage, will not be able to generate and redistribute sufficient productivity gains to guarantee their populations decent living conditions.

We thus understand that for states—if we accept that they are social institutions with their own objectives and a desire to perpetuate themselves autonomously—the race for artificial intelligence amounts to a race for the survival of the regimes themselves.

Is artificial intelligence a weapon of mass destruction? What we can say, at the very least, is that by attributing demonic powers to artificial intelligence, we may no longer be far from a self-fulfilling prophecy—what psychology calls the Golem effect—yet another story of an artificial creature that risks escaping the control of its creator.

Fantasies should not be dismissed, for sometimes their effects are more real than their causes. One last point I would like to raise to conclude this speech is the political vision held by the leaders of the major firms producing artificial intelligence technologies. We have only recently begun to pay attention to it, but with Musk, we see a textbook example of a reactionary drift, masculinist, racist, a kind of blend of ultranationalism and libertarianism that is spreading through parts of the American tech industry.

We tend to reduce this issue to two main components: either the far-right tendencies of Silicon Valley, or, in a more psychological interpretation, the delusional structure of certain individuals. Yet, it is crucial to take seriously what people like Elon Musk, Peter Thiel, Mark Zuckerberg, etc., are saying. They are talking about AI, not just from a technological or economic standpoint. Take someone like Ray Kurzweil, for instance. He is the Director of Engineering at Google, co-founder of Singularity University, an incubator for startups, a think tank, and an educational center. In his well-known book "The Singularity is Near", published in 2005, Kurzweil predicts that, thanks to the

progress of artificial intelligence, machines will surpass human intelligence by the year 2045.

This is what he calls the dawn of technological singularity. He explains that the imminent arrival of this technological singularity—what could also be called strong AI—poses the risk of humans becoming slaves to machines, much like in *The Matrix*. To avoid being reduced to servitude, he argues, humans must adapt and use technology to enhance themselves and thus remain competitive against these new rivals: the machines.

I still follow his perspective. His idea—shared by Musk and others—is to enhance human intellectual and physical capacities, or even to transfer consciousness itself into a machine. However, what is interesting and important to keep in mind is that these transhumanist prospects are envisioned as being reserved for a small plutocratic elite. The future imagined by these tech leaders is one of a new social order, a kind of oligarchic timocracy in which they would sit at the top, while the rest of the population would be condemned to wander aimlessly and without resources, unable to compete with the productive power of artificial intelligence.



It is, in essence, the dream, the fantasy of a return to a society of orders, a strictly aristocratic society, governed by the aristoi, the “best”, in this case meaning those who have achieved intellectual and physical superiority over both humans and machines.

It is, in essence, the dream, the fantasy of a return to a society of orders, a strictly aristocratic society, governed by the *aristoi*, the “best”, in this case meaning those who have achieved intellectual and physical superiority over both humans and machines. This may well be the point where we can truly see artificial intelligence as a weapon of destruction: the moment it becomes the technological foundation of a political project aimed at dissolving democracy.

When we see this new billionaire, Minister of Government Efficiency in the world's leading power, shamelessly performing a Nazi salute,

calling for the annexation of democratic countries, and championing apartheid, then perhaps it is indeed not so far-fetched to think of artificial intelligence as a potential weapon of mass destruction. Thank you.

Yannick MENECEUR: To conclude this conference, among all of today's contributions, one sentence from Antoinette ROUVROY particularly caught my attention: Justice is something dialectical, always something for the future. You will not find it in the data. But you will most certainly find injustice in the data. ■

References

Books

- BARTHES Roland, *Le Neutre. Cours au Collège de France (1978)*, Seuil, 2023.
- BERGSON Henri, *La Pensée et le Mouvant* (1934), PUF, 2013.
- GANASCIA Jean-Gabriel, *L'Âme-machine. Les enjeux de l'intelligence artificielle*, Seuil, 1990.
- GANASCIA Jean-Gabriel, *Le Mythe de la Singularité. Faut-il craindre l'intelligence artificielle ?*, Seuil, 2017.
- GANASCIA Jean-Gabriel, *L'I.A. expliquée aux humains*, Seuil, 2024.
- GARAPON Antoine et LASSÈGUE Jean, *Justice digitale*, PUF, 2018.
- HELER Joseph, *Catch 22* (1961), Grasset, 2004.
- KANT Emmanuel, *La Critique de la raison pure* (1781), PUF, 2012.
- KURZWEIL Ray, *The Singularity is Near*, Viking, 2005.
- MENECEUR Yannick, *IA générative et professionnels du droit. Comprendre et s'appropriier la langue des probables*, LexisNexis, 2024.
- O'NEIL Cathy, *Weapons of Math Destruction*, Crown Publishing Group, 2016.
- RANCIÈRE Jacques, *Le Partage du sensible : Esthétique et politique*, La Fabrique, 2000.
- SONTAG KOENIG Sophie, *Technologies de l'information et de la communication et défense pénale*, Mare & Martin, 2016.

Articles, proceedings and research works

- BIGELOW Julian, ROSENBLUETH Arturo et WIENER Norbert, « Behavior, Purpose and Teleology », *Philosophy of Science*, Volume 10, Issue 1, 1943.
- LEGRAND Emmanuelle et POPA-FABRE Murielle, « IA Générative et décision du juge », dans *Impact du numérique sur la Justice*, Volume 2, Institut Robert Badinter, 2024.
- *Les nouveaux défis de la cyberjustice : entre information, régulation et démocratie*, Actes de la Conférence Cyberjustice Europe 2023, Institut Robert Badinter, 2024.
- McCULLOCH Warren et PITTS Walter, « A Logical Calculus of the Ideas Immanent in Nervous Activity », *Bulletin of Mathematical Biophysics*, Volume 5, 1943.
- SONTAG KOENIG Sophie, « Dispositifs audiovisuels, information juridique et judiciaire, règlement en ligne des différends », dans *Impact du numérique sur la Justice*, Volume 1, Institut Robert Badinter, 2024.
- THIBOUT Charles, « Google : une entreprise d'influence ? Indigénisation, normalisation et politisation d'une multinationale du numérique en France. », 2024.

Videography

- Conférence Cyberjustice Europe 2023 on YouTube : <http://bit.ly/43z2tbb>
- Conférence Cyberjustice Europe 2016 on YouTube : <http://bit.ly/43z3a4h>
- *iHuman: L'intelligence artificielle et nous*, Tonje Hessen Schei, 2019.



Conference program

Decoding AI in 2025: Insights into a technology on its way of normalization

January 23, 2025, 9h00-17h00

Centre de Culture Numérique,
Strasbourg, France

9h00

Welcome

9h15

Welcome address by the organizers

- Jeanne-Marie TUFFERY-ANDRIEU, Dean of the University of Strasbourg's Faculty of Law, Political Science and Management
- Yannick MENECEUR, Magistrate and associate lecturer at the University of Strasbourg's Faculty of Law, Political Science and Management
- Harold ÉPINEUSE, Deputy director of the "Institut des études et de la recherche sur le droit et la justice" (IERDJ)

9h30

Opening conference:

But where is AI going?

- Understanding the foundations and ethical implications of artificial intelligence
 - Taking a step back from preconceived ideas about artificial intelligence
 - A critical reflection on the motivations of major technological players is in order
- Jean-Gabriel GANASCIA, Professor of Computer Science at Sorbonne University's Faculty of Science, senior member of the "Institut Universitaire de France" and Chairman of the CNRS Ethics Committee

10h00

Fireside chat:

Speaking the language of generative AI

- What are the latest advances in generative artificial intelligence?
 - Understanding the conceptual foundations of artificial intelligence models to optimize their use and guarantee their governance
 - Automation must be achieved by giving it meaning
- Interviews conducted by: Yannick MENECEUR, Magistrate and associate lecturer at the University of Strasbourg's Faculty of Law, Political Science and Management
 - Murielle POPA-FABRE, Expert at the Council of Europe on issues of artificial intelligence regulation and data protection
 - Thomas LAMPERT, Professor of Artificial Intelligence and Data Science at the University of Strasbourg

10h30

Health break

10h50

Screening of a short excerpt from the documentary "iHuman" (ARTE)

11h00**Round table:****Will AI transform freedom into destiny?**

- Analyze the societal and political implications of artificial intelligence on our individual freedom
 - Assess the dangers of algorithmic determinism, which could reduce the field of possibilities
 - Promote collective, democratic reflection on the integration of artificial intelligence into our societies
- Moderation: Harold ÉPINEUSE, Deputy director of the IERDJ
 - Maxime DERIAN, Doctor of digital socio-anthropology
 - Antoinette ROUVROY, FNRS researcher at the "Centre de Recherche en Information Droit et Société" (CRIDS) at the University of Namur
 - Bilel BENBOUZID, Senior lecturer at the University Gustave Eiffel, "Laboratoire Interdisciplinaire Sciences Innovations Sociétés" (LSIS)
 - Sabine VAN HAECKE LEPIC, Lawyer and Doctor of law

12h30**Lunch break****14h00****Round table:****AI and legal professionals, a great misunderstanding?**

- Deciphering expectations and issues linked to the use of AI in the law: greater efficiency, to be objectivized
 - Audacity, transparency and caution appear central in the use of artificial intelligence systems in the legal field
 - Concrete projects already in the legal professions
- Moderation: Sophie SONTAG KOENIG, Senior lecturer at the University of Paris-Nanterre
 - Aurélie KLEIN, Lawyer and Associate Lecturer at the University of Strasbourg's Faculty of Law, Political Science and Management
 - Frédérique BERROD, Professor of EU law at Sciences Po Strasbourg
 - Pascal ALIX, Lawyer
 - Claire STRUGALA, Judicial magistrate and Doctor of law
 - Marc CLÉMENT, Administrative judge

15h30**Closing conference:****AI, a weapon of mass destruction?**

- Artificial intelligence is a marketing fantasy, its origin is not really aimed at reproducing human intelligence but at simulating it
 - The historical use of artificial intelligence from the Second World War to its current deployment in autonomous weapons systems is military
 - The social and political consequences of automation are to be apprehended from the angle of massive job destruction and the instability of political regimes
- Charles THIBOUT, Doctor of political science, member of the "Centre européen de sociologie et de science politique" (CNRS, EHESS, Paris 1 Panthéon-Sorbonne)

16h00**End of the online broadcasting****16h00-16h30****Discussion between the audience and speakers**

- Joint moderation by Sophie SONTAG KOENIG, Harold ÉPINEUSE, Yannick MENECEUR

Directrice de la publication : Valérie SAGANT
Auteurs : sous la dir. Harold ÉPINEUSE, Yannick MENECEUR
Re transcription : Nathalie NIELSON
Suivi d'édition : Xavier CAYON et Léa DELION
Réalisation : Isabelle CHEMIN
Secrétariat de rédaction : Coralie MANANT
Photos : Vidéo de la conférence Décoder l'IA <http://bit.ly/3JswtyJ>

DECODING AI IN 2025 INSIGHTS INTO A TECHNOLOGY ON ITS WAY OF NORMALIZATION

This public conference organized early 2025, outcome of a collaboration between the “Institut des études et de la recherche sur le droit et la justice” (now the Robert Badinter Institute) and the University of Strasbourg, marks the first step in a series of events aimed at “decoding AI”, proceedings of which will be published in several volumes.

In this first volume, intended for a wide audience of practitioners, researchers, and students, we examine the challenges posed by the widespread use of digital technologies known as “artificial intelligence” in their technical, social, and legal dimensions through:

- a review of some basic technical concepts in simple terms,
- an overview of the political and societal challenges of AI early 2025,
- an up-to-date and practical assessment of the use of AI in the legal and judicial professions.

Based on the interdisciplinary discussions that took place throughout the conference, these proceedings share some of the essential knowledge needed to develop a common language of AI among lawyers, computer scientists, philosophers, sociologists, and political scientists.



Postal Mail

Institut Robert Badinter
Ministère de la Justice
13, place Vendôme
75042 Paris Cedex 01

Visit our premises

47 bis, rue des Vinaigriers
75010 Paris

contact@institutrobertbadinter.fr
institutrobertbadinter.fr



INSTITUT ROBERT BADINTER Études et recherches sur le droit et la justice

The Robert Badinter Institute (formerly the Institute for Studies and Research on Law and Justice - IERDJ) aims to foster knowledge and discussion on law and justice by initiating original and forward-looking thinking, funding and supporting research projects, and widely sharing knowledge on standards, regulation, and the functioning of justice across all scientific disciplines.

A public interest group created in 2022, the Robert Badinter Institute is the result of a merger between the Institute for Advanced Studies on Justice (“Institut des hautes études sur la justice”) and the Law and Justice Research Mission (“Mission de recherche droit et justice”), two entities that have long been internationally recognized.