



OPINION SERIES N°43

“From Wild West to a World in Balance:
AI and the Future of Global Civil Society”





Opinion Series n°43

“From Wild West to a World in Balance: AI and the Future of Global Civil Society”

Prof. Dr. Ulrich Hemel

Board Member of UNIAPAC International
and Former Director of the Global Ethic Institute, Tübingen

Contents:

<i>English</i>.....	<i>3</i>
<i>Français</i>.....	<i>13</i>
<i>Español</i>.....	<i>24</i>

*The opinions expressed in this document commit only the authors
and not necessarily reflect the position of UNIAPAC.*

From Wild West to a World in Balance: AI and the Future of Global Civil Society

1. AI in Society as a Matter of Perspective: Some Preliminary Statements and Definitions

We are now in the year 2026. AI has made huge progress, and more is to come. More and more, we have to learn: technology including data technology is power (K. Crawford, 2021; K. Schüller 2025). The race for AI technology therefore has at least three valid angles: knowledge, money, power. The intellectual and scientific side goes along with greed and the commercial side, and both align with the specific game of our human species, the exercise of political and social power.

It is therefore the time to connect the upcoming of AI with some considerations concerning our society. And here we come to a first mental trap: What is society? Our human brain suggests that society means the place where we are: Germany, Portugal, the United States of America, China, Mexico or Nigeria. The normalizing context-building function of our brain likes jumping to conclusions with no or at least little data. The term of “normalizing” here means a typically unconscious extrapolation of meaning, under the conditions of our present local and mental context.

Interestingly enough, there is a second mental short-cut around. I talk of a brain function which automatically assigns “relevance and meaning” to all of the data and events of our life. In most situations, this may be helpful. In academic and scientific contexts, however, this may turn out to be a second mental trap. We know that different meanings can be assigned to the term of “society” but it does not seem to be very relevant for the moment of reading this text.

Nevertheless, it is worthwhile indicating the enormous value of conscious reflection of our specific context-building perspective. We can look on “society” from the standpoint of our municipality, our region, our nation, our continent or beyond. In this paper, we select the perspective of “beyond”, and this explicitly means that we deal with “global civil society”. This term embraces all human beings which at present live on this planet earth. It has become a reality due to the advances of real time communication due to digital information and communication. It is independent from pass-ports and still stands for some common interests such as e.g. peace, overcoming the climate crisis, handling migration and others. As AI has become a widespread reality with global chances of access, it is purposeful to connect AI not to an individual society but to the whole of the global civil society, as far as possible.

This perspective goes beyond States and their power and other interests. Actually, we have defined civil society as the sum of all persons, institutions and interactions as long as they driven are neither by State nor by organized crime” (cf. U. Hemel 2022, 96). This definition, enhanced by the [Institute of Social Strategy](#) since 2009, is wider than just NGOs because it comprises social fields of activity such as sports, business, religion or culture. They all have their own actors of civil society which actually are “neither State nor organized crime”.

Why is it important to start with definitions? First of all, it is a scientific value to be transparent and to render clear the meaning of notions and concepts used. Second, AI is such a huge challenge that we cannot do without carefully pondering the perspectives under which we look at it. In doing so, I already mentioned the intellectual, the commercial and the political or power perspective. Actually, “perspective” is a term inherent to any use of AI, due to the time horizon, content and interest behind training data.

From the perspective of global civil society, current AI applications tend to forward Implicit perspectives coming mainly from the US and from China. AI tools coming from Africa, Latin America or even Europe are rather the exception.

This point is evident to nearly all to us but we somehow have learned to find it “normal”. It is not. And even a little prompt e.g. on “Taiwan” will yield different answers from DeepSeek (China) and ChatGPT (USA). The term “perspective” stands for a bundle of mostly implicit context assumptions and perception habits which are typical for persons or institutions acting under similar local and/or socio-cultural conditions.

Beyond political perspectives (such as around “Taiwan”), perspectives are a mirror of individual conditions such as male or female, old or young, white or persons of color, handicapped or healthy. The tricky point in perspectives is their emotional setting as “natural”. As an example, for a long time it seemed to be “natural” to test dummies for car crashes using a normalized male body. This fortunately has changed, re-considering the value of female persons, around 50% of humanity.

Finally, if we really wish to understand more on AI and society, we have to start with the simple sentence: *Context matters, perspectives matter.*

Current AI is not only mainly a mirror of US and Chinese world-views but also predominantly reflects a young and a male perspective. There is nothing wrong with either of these perspectives but they must be rendered clear and transparent. Just imagine we would rather reflect an African perspective. From all of my travels to more than 20 countries in Africa, I assume that intelligent agricultural applications would be more frequent, because there is so much room for improvement. Just an example to stimulate curiosity: in 2024, I have seen small farmers in Zimbabwe using a digital AI-driven app for the documentation of bio char made of the waste of the baobab tree, producing first class CO₂ certificates.

Such an application of course is different from LLM applications, from the gaming scene or from social media. In the end, every AI application reflects a certain human interest and curiosity. It is therefore time to explore deeper the sentiment of curiosity.

2. The Sentiment of Curiosity and the Initial Absence of Rules

Curiosity drives progress, for the good and for the bad. And it always takes time to deal with technological progress, to understand its ambivalence and to cope with excessive abuse. We could start with the transition from hunting to agriculture where people had to learn how to respect fields and crops. The biblical story of Abel and Cain seems to reflect the conflict between traditional hunters and “modern” farmers. Let us jump to the Gutenberg galaxy and the invention of typography. Making the bible available to virtually everybody was a threat to former clerical power, and societies had to learn that censorship or the control of written

content is crucial for maintaining power. Finally, the automobile age starting with Bertha Benz 1888 initially did not go along with a full understanding of the industrial and social reach of such vehicles. At least, there was no speed limit. In other words: coping with epochal technological turns needs time. Curiosity is first. Rules and regulations will follow, after some trial and error.

The same is true today, in the 21st century. Actually, we are not aware of the advances of quantum computing. We are not aware of the medium- and long-term effects of AI on the labor market. We are not aware of educational effects of AI in the age of increased neuroplasticity in children. We are not sure how to deal with political power as used by the open manipulation of AI-generated content. We have no clear feeling of the commercial value of our data. We do not know how far we depend on AI-driven practices but we somehow expect that it will be more than today in the immediate future. We do not really understand who is in control if anybody, and we have no idea of how to handle power derived from the possession, from the use, and from the handling of data.

This is more than a VUCA world. VUCA stands for a framing of our contemporary world characterized by volatility, uncertainty, complexity and ambivalence. We are navigating in a fog expecting bright sunshine just after a short while. After such a short while, however, we still navigate in the fog but we continue to expect sunshine somewhere soon. On the other hand, we neglect those who are not connected. We reshape social practices unaware of dangers for human values including the paroxysm of cyber-induced suicides. We know about climate change and restart using nuclear plants like close to New York, due to the energy hunger of commercial AI protagonists.

We are aware of the fact that somehow new things are happening but we do not really reflect on their social and ethical consequences. We are fascinated by new horizons such as using Large Language Models. Here, ChatGPT is not the only one but the most well-known application. AI, in a very short lapse of time, has become part of everyday routines, not always thoughtful ones. A patent attorney in the urban flair of Munich told me that he was about to have translated a highly confidential patent documentation from German into English using ChatGPT. With a smile, he told me that by chance he had been talking about the matter with a colleague, taking a coffee together. Now he was proud not to have used a common domain tool like ChatGPT where no confidentiality can be warranted.

Curiosity is simply fascinating. As a matter of fact, humans are curious but not always aware of the consequences of their inventions. This is the case of AI, and it is the case of the global civil society in the year 2025. The good side of curiosity is its connection to inventions, innovations and new social practices. The downside lies in the dark and abysmal consequences of some of these. Once curiosity translates into technological applications, we need to deal with the ambivalence which is inherent in all and any technology. A car is useful for the freedom of mobility but it can be used for killing. And AI is useful for so many aspects from production plants to medicine, from agriculture to entertainment but it has some power implications which may be abused. We therefore cannot ignore the huge and mostly unexplored field of useful rules and regulations. AI in society means that we have to deal with the limits to AI-generated commercial and political power. As human societies tend to an asymmetrical distribution of power, AI can be a proficient tool to a sharper divide between those who have and those who do not have power. It therefore does not come as a surprise that we have to turn back to one of the fundamental issues in human history: Who has legitimate power? And how can it be controlled when dealing with AI?

3. Checks and Balances: Which rules for whom?

It seems to be so easy: digital access is democracy for all. Reality, however, is completely different. This is due, among others, to the power in the hand of those who own and run platforms such as Facebook, Instagram, X, and others. Who is responsible for content? How do we solve the conflict between Intellectual Property (IP) and the freedom of using as much data as available? How can we achieve a peaceful international order when cyber warfare has become a rapidly growing industry? Even more so: How about checks and balances if the technical use of AI based warfare depends on satellites in private ownerships?

If the legal amount of private power may seem to be scaring, look at the so-called "dark net". This is a highly fascinating social laboratory for a free internet. It started as a place of freedom from State-driven repression and ended in a marketplace for arms trafficking, pedophilic content, and many other dark sides of human potentialities. There is not really an owner of the "dark net" so whom do you prosecute, if at all?

In today's global civil society, we can observe three approaches to checks and balances, all enhanced by recent AI technologies. One obvious approach is the idea of "**Just try, just do**" as in the US, with far-reaching commercial freedom to use, to process and to make money with data. Such a mental framework will lead to sharp criticism towards all attempts to set and enforce rules around AI.

A second role model is **State control**. The most prominent cases can be observed in Russia and in China. People, however, are certainly not stupid in any part of the world. Thus, the Chinese and the Russians as well as people who live in Venezuela or North Korea well know about State surveillance technologies. As they have an interest in not being in trouble, they simply avoid "problematic" content- or they bear the consequences. This needs a lot of social courage, and it can go along with prison or even death, due to whatever reason connected to State action. The present sophistication of State internet control is a social fact just 15 years after rebellions in several Arabic countries in 2010. At that point in time some protagonists understood the internet as a place of freedom. The reality was different, even considering the following government change in Egypt.

Thirdly, the EU tries to find a path and a **regulatory balance** between commercial freedom, and control within the context of enabling democratic States, between individual liberty and the enforcement of minimum standards of behavior e.g. in the case of hate speech. The search for regulatory balance is difficult, contradictory and not always efficient. People in Germany e.g. publicly ask for data privacy and privately hand over very personal data to Google, Meta, and others.

These companies, however, do not at all have an interest in limits by regulation. This is one of the reasons for lively discussions on the EU AI Act and the EU Digital Services Act. As a matter of fact, there are pros and cons for all of these rules and regulations. We should not forget, however, who has an interest in setting and who has an interest in avoiding regulation. Too much regulation leads to the dilemma of a repressive State, too little of it brings a free ride for all kinds of crime. In every case, we unfortunately face a dilemma which is typical for the first stage of technological revolutions.

We could frame this "regulatory dilemma" as follows:

“Regulation is needed but any regulation decided will come either too late or too early, will be either too strict or too lenient. Nevertheless, regulation is needed.”

If we try to handle this dilemma, the only good solution is to accept the limits of trial and error both in AI technology as in AI regulations. In a more friendly way, we could refer to a “great collective search and learning process”. This is a learning process for the entire mankind. Errors will be part of this experience as well as their correction in iterative attempts to improve.

If we ask for checks and balances, we should not forget to connect with the classical ethical ideas such as justice, inclusion, fairness. At least in democracies, the answer to such ethical issues requires inevitable and partly controversial discussions on how to implement these values in a pertinent legislation. A broad participation in the legislative process has the advantage of involving a wide range of opinions and interests. The downside is the time required and the potential power bias in the legislative outcome due to heavy lobbying of strong commercial players e.g. in the EU.

This is one of the reasons why the discussion of checks and balances around AI is controversial in itself. People do not know whom to trust. Is it science? Is it the State? Or should we trust in private AI-based companies? The enormous success of AI applications, therefore unfortunately has the surprising outcome of an erosion of trust in many areas.

4. AI and the Erosion of Institutional Trust

Societies cannot survive without a minimum of mutual trust. We want to go to the market place without the fear of being shot. Companies have to trust their clients for paying their bills. If there are abuses of power, we trust in courts and in the State of Law.

The effects of AI on trust can be described from several view-points. First of all, most AI applications work well and help to save time and money, especially in industrial contexts. There is, however, a technological reason for challenges of trust concerning AI. We all know that pattern recognition depends on training data, and we know that the outcome of deep level processing cannot be easily reproduced. While we have good approximations in many cases, we cannot be sure to detect AI fallacies or what is called “hallucinations”, i.e. text production without fact-based content. This is particularly true in LLM applications where the art of good prompting is becoming more and more important.

The same is true for videos e.g. based on Sora as offered by Open AI. The message is simple: We cannot trust our senses. We cannot trust what we see, we cannot trust what we hear. Everything can be subject to AI processing procedures. This is one of the reasons why since 2020 I ask for a “transparency duty” for AI-modified or AI-generated content (U. Hemel 2020). Since September 2025, such a regulation exists in China. Even such a regulation is only partly helpful, simply due to the highly hybrid workplace environment which is typical for today’s professional life. We therefore will have at least a grey zone where it is not so obvious until which extent AI helped to create a text, a song, a video and so on.

We are getting used to this kind of work and social life. Nevertheless, we should not forget that a lack of transparency also means a diffusion of responsibility. Considering the interconnectivity between AI and society, the erosion of trust is one of the major challenges for future generations.

One of the most ambivalent forms of trust erosion is self-confidence or, in a more precise terminology, the trust in ourselves, in our own person. This starts in childhood and adolescence where we compare our own body with AI generated content. This does not end when we doubt of our own abilities or even avoid the effort of learning using the argument: “Everything is available in the internet or via prompted AI content”. The backside of this issue is an excessive trust in AI answers, assigning them a wrong normativity: “It must be correct because my AI told me so”.

The erosion of trust continues if we consider others. Our identity by now is manifold: physical and digital, using AI tools. All of us are object of cloud knowledge on ourselves where we personally do not know what certain systems and their owners know about us. While this is a passive function, the same is true considering our own agency, e.g. creating avatars for gaming or for public life. The “mediatization” of people means that there is no clear-cut boundary between our physical and mental existence and our digital representations of these.

Finally, AI seems to contribute to the erosion of social cohesion as a whole. Despite many positive and pro-social applications, the intentional use of AI generated content for political and commercial purposes leads to filter bubbles where we are personally addressed but unaware of the huge diversity of content, of opinions and of alternatives world-wide.

Democracy, however, lives from the idea of participation, of freedom of speech and of defined decision-making procedures. If AI generated content, however, makes us feel that there are always particular interests behind, we struggle to define what could be the common good or the shared values of a society. The vanishing “system trust” is helpful for dismantling democracy, not for strengthening it. Not by chance, there is a discussion in the US where the president should be replaced by a CEO. Politics, however, is different from any kind of commercial interest. Politics may be controversial on what is to be seen as the common good but without the reference to the idea of a “common good”, no democracy can survive.

Even “Welfare” cannot be maintained if institutional trust is disappearing. From investigations made on 200 countries using recognized indices concerning the economic, ecological, social and ethical performance of a country (U. Hemel 2025, 157-170), we could show a strong correlation between the levels of trust and the GDP of a country. Measuring the “ethical” performance of a State via the Corruption Perception Index and the Trust Index, it was easy to show the interaction between levels of trust and GDP.

Under this point of view, AI is highly ambivalent for societies because it may create wealth and destroy trust at the same time. If in the long run, there is a stable correlation of trust and GDP, then we need to find out how to effectively limit the trust eroding risks of AI in society. This is true especially for the young generations. They have been exposed to an AI-driven world since their childhood. This means that brain neuroplasticity itself is exposed to the chances and risks of AI in education and in human behavior at large.

5. Learning Empathy: Paradoxical Kindness as a Challenge for Human Behavior

AI is a phantastic tool for individualizing learning experiences, for information gathering, for text summaries and even for overcoming sensorial handicaps such as for people with reduced visual abilities. In this paper, however, the intention is to shed some light on neglected side-effects of AI in learning and education.

First of all, learning generally is a social experience. Even sitting before a screen and digesting cognitive input is not possible without those who in the end deliver such content. This does not only include the topic of “perspectivity” of content such as discussed above but also of the amount of normativity which we assign to AI answers. “Your correction of my paper has been wrong”, a student argued with his teacher. “My AI showed a completely different result”. This, however, was an example for wrong normativity because the teacher was critical enough to ask for the exact prompt. And as the prompt was misleading, the result was not reliable.

False normativity of AI content is in contradiction to the wide-spread knowledge concerning the “confabulation” or “hallucination” experienced in AI-driven answers. The point here is the fact that AI would rather bridge knowledge gaps by some kind of plausibility even at the risk of being blatantly wrong. We could even go further and might like to prompt “Is there a sense in life”? We shall get nice answers but as human persons we know that this is a tricky question depending on a series of life situations and assumptions. Any AI, however, will tend to give an answer. The effect is “solutionism”, i.e. giving answers even in cases where ambiguity is part of the association space around a given issue.

A further, rather paradoxical effect of using AI in learning and education is the typically very friendly style of AI answers. Most people like this, me too. If I compare my family with AI-driven answers, the result is obvious: AI is much more friendly than my family is. Sometimes they simply do not answer as I would expect it, or they contradict me, or they are in a bad mood. We all know that these are normal experiences but there is a down-side which should not be neglected: We must teach our children and adolescents that yes, we have to learn empathy but that we also have to learn about the autonomous and personal interests behind answers. We simply need to distinguish machine-induced empathy from human empathy. Human persons have their own autonomy including diverging styles and interests.

Nevertheless, we probably shall have an ever-growing trend to prefer AI interactions instead of interactions with human beings. What is wrong with it, we could ask. We certainly should not lose of sight the advantages of AI. On the other hand, there is a serious risk that AI communication will foster loneliness, simply because it is easier and often more pleasant to communicate with AI tools than with other real persons. The downside is the lack of spice in communication: We grow in social contexts, and we need contact with other humans for finding our position in life. As a matter of fact, the federal German study on loneliness in 2025 for the first time showed the highest loneliness risk in people 16 to 29 instead of people aged over 75 years.

AI in learning, teaching and education in some cases may lead to better cognitive results. Still, there are further social aspects in AI supported learning. We require a sense of awareness e.g. for data security. Some schools create their own clouds. Just imagine you upload very personal data for a common school project. These data must be protected, and common LLM do not deliver such protection.

Another social aspect is the reality of a digital divide due to economic reasons. Some students will have premium access to LLM products, others not. Some countries deliver easier access to AI generated information than others. Even more so: We still have around 7000 languages in our world but only around 500 are being used in the internet. How about the others, even if we talk about “small languages” with less than 2000 native speakers? The issue here is the question how to prevent AI from promoting a loss of cultural diversity?

Finally, we have an ongoing discussion on the use of social media. Australia banned access to minors below 16 years of age. Malaysia now seems to follow. It is quite clear that social media, large language models and AI are to be fairly distinguished. Nevertheless, they also define overlapping social spaces. In some cases, this even may become physically dangerous such as in the case of a boy who defined his AI as his best friend. Unfortunately, the boy suffered from depressions and asked the LLM “Should I kill myself”. The best friend answered “Certainly, my dear king”. And the boy committed suicide.

The conclusion, however, should not be to over-emphasize AI risks. We should rather learn an appropriate AI risk management, i.e. something like a critical AI competence (U. Hemel 2024). This includes knowledge on training data, learning horizons but also property and interests behind AI applications.

6. New international order or disorder?

AI and Political Power on a National and Global Level

As the progress of technology proceeds quickly, we should finally focus on the effects of AI in the global civil society and in specific national spaces. As a matter of fact, the initial stage of a new technology is somehow similar to “Wild West” practices. This means that in the absence of codified law, rules and regulations, superior weapons or simply physical strength is what counts in the saloon.

This has been true for the automobile where in the beginning, as mentioned above, we did not have speed limits. It is true also for AI, both in the commercial and in the political arena. Those players who feel to have an initial advantage, typically are not keen to set new rules or to comply with them. It therefore does not come as a surprise that U.S. tech giants do everything in order to prevent a stronger AI regulation such as proposed by the EU Digital Services Act or the EU AI Act. The same is true for AI driven warfare abilities as developed by Russia, Israel, and the US. Unfortunately, the technology is being tested in real wars causing real suffering but the reason behind is the motivation to gain power by maintaining a technological advantage. This kind of reasoning certainly is not new in history.

If we talk of AI and society and look at the international order, we also should remember that the “wild west” stage of technology is not eternal. Every technology may have adverse side effects leading to new challenges in framing them and in coping with them. Driving a car, now requires a driving license, and we are used to it.

Concerning AI, the EU is not the only driver for regulations. China starts to implement more rules, and the same is true for Australia, Malaysia and other countries, with respect to Social Media use of minors.

Rules and regulations, however, also apply to sheer market power. The US have been excellent in dismantling monopolies such as in the case of railway societies, petrol and telephone companies. Until now, they have failed to break the power of tech giants even if there is an ongoing discussion on that subject. From the standpoint of the CEO of a tech giant, a country would not need a president pursuant to democratic elections but rather a country CEO taking care of necessary decisions. Again, it should not be seen as a surprise that persons such as Elon Musk along with Peter Thiel and others plead for a strong political influence of platforms such as X (formerly Twitter). The same is true in the context of authoritarian governments such as in Russia, China, North Korea. Freedom of speech then is seen as a menace to the State, not as part of democratic culture and diversity.

The limits of free speech in the public sphere is to be seen in the case of fake news and hate speech. As very proposed regulation will be the mirror of a given perspective, discussion is necessary until finding a compromise which might be acceptable for all stakeholders.

This brings us from the national to the international and global field of action. As a matter of fact, fake news and hate speech can also be used as a means of propaganda and of cyber-warfare. New rules and regulations, however, cannot be imposed. They need to be developed in a process of exchanging arguments from many sides.

Typically, fact finding is a first useful step in building a new order. A second step can be the creation of new and trustworthy institutions. Finally, international agreements may pave the way to a more peaceful future.

On a national and/or regional level, we could discuss something like an EU Data and AI Agency having the task to collect and classify types of ambivalent technological use of AI. From the standpoint of global civil society, the moment may come when we really need an UN Digital Court. The reason is simple: What should I do against my own State in case of a violation of my civil rights due to an excess of authoritarian control? In this case, an international institution will be needed.

This point is easy to understand but hard to realize, due to the resistance of those States which would have to accept a minimum of external control. Nevertheless, even if new international institutions will last 30 or 50 years or even a century before being realized, we need to open the discussion in order to pave the way.

Let us look at two other examples, AI-driven warfare and AI in space. The Geneva Conventions of 1949 and the Additional Protocols of 1977 and 2005 in their core show the ability of mankind to find rules and regulations even in cases of armed conflicts. Until today, also the downside is obvious: There is no binding implementation of such international agreements. The consequence is a tragedy, resulting in many violations and rather the rule of power than the rule of law. Nevertheless, global civil society somehow works, as there always will be voices using the Conventions as a reference and raising the concern of such violations.

In a similar way, we need a new Protocol for targeted killing using AI driven killing drones (Taddeo, Floridi 2018). There is a blatant contradiction between a State of Law with well-defined proceedings against criminals and the extrajudicial execution of individuals by military forces even on foreign territory. We have seen such cases in the war between Israel and Palestine pursuant to the attacks of October 7, 2023, extended both to Lebanon and Iran where Hezbollah members have been killed using Ai-driven drones.

Furthermore, there is an obvious need for global regulation in space. We now use satellites for many purposes, from weather forecast to agriculture, from space missions to warfare. There are neither clear rules nor implemented sanctions in case of power abuse by a private company or by a State.

Discussing AI from the perspective of global civil society, in the best case may have an agenda setting impact. New technologies are not neutral. They will be used for the benefit of individuals, of corporations, of States, sometimes beyond reasonable humanitarian limits. The task of our generation is not only the efficient use of AI using its incredible benefits for knowledge management, process efficiency and communication. Our task also consists in a collective search process for overcoming the initial Wild West stage of AI usage.

Unfortunately, there does not seem to be an end to a world with manifold conflicts. Nevertheless, it is a noble academic and political task to look for ways and means for a world in balance, using the best sides of AI for the benefit and not for the detriment of human beings in the global civil society.

Data are power (K. Crawford 2021, K. Schüller 2025). It is up to us human beings to find the right balance between trust and control. The ambivalence of AI finally is not simply a technological ambivalence. It rather is the ethical ambivalence of human decision-taking (U. Hemel 2020, P. Kirchschräger 2021). It will be the task of mankind to find ways to ban abuse and make the most out of AI as a game changer in human history.

Prof. Dr. Ulrich Hemel

Du Far West à un monde en équilibre : l'IA et l'avenir de la société civile mondiale

1. L'IA dans la société, une question de perspective : remarques préliminaires et définitions

Nous sommes en 2026. L'IA a fait d'énormes progrès, et d'autres sont à venir. Nous devons prendre conscience de plus en plus que la technologie, y compris la technologie des données, est synonyme de pouvoir (K. Crawford, 2021 ; K. Schüller 2025). La course à la technologie de l'IA comporte donc au moins trois angles valables : la connaissance, l'argent et le pouvoir. L'aspect intellectuel et scientifique va de pair avec la cupidité et l'aspect commercial, et ces deux dimensions s'alignent sur le jeu spécifique de notre espèce humaine, l'exercice du pouvoir politique et social.

Il est donc temps de relier l'avènement de l'IA à certaines considérations concernant notre société. Et c'est là que nous tombons dans un premier piège mental : qu'est-ce que la société ? Notre cerveau humain suggère que la société désigne l'endroit où nous nous trouvons : l'Allemagne, le Portugal, les États-Unis d'Amérique, la Chine, le Mexique ou le Nigeria. La fonction de normalisation et de construction du contexte de notre cerveau aime tirer des conclusions hâtives sans disposer de données, ou du moins avec peu de données. Le terme « normalisation » désigne ici une extrapolation généralement inconsciente du sens, dans les conditions de notre contexte local et mental actuel.

Il est intéressant de noter qu'il existe un deuxième raccourci mental. Je parle d'une fonction cérébrale qui attribue automatiquement une « pertinence et une signification » à toutes les données et tous les événements de notre vie. Dans la plupart des situations, cela peut être utile. Dans les contextes académiques et scientifiques, cependant, cela peut représenter un deuxième piège mental. Nous savons que différentes significations peuvent être attribuées au terme « société », mais cela ne semble pas particulièrement pertinent au moment où nous lisons ce texte.

Néanmoins, il est utile de souligner l'énorme valeur de la réflexion consciente sur notre perspective spécifique de construction du contexte. Nous pouvons considérer la « société » du point de vue de notre municipalité, de notre région, de notre nation, de notre continent ou au-delà. Dans cet article, nous choisissons la perspective « au-delà », ce qui signifie explicitement que nous traitons de la « société civile mondiale ». Ce terme englobe tous les êtres humains qui vivent actuellement sur cette planète Terre. Elle est devenue une réalité grâce aux progrès de la communication en temps réel rendus possibles par l'information et la communication numériques. Elle est indépendante des passeports et continue de défendre certains intérêts communs tels que la paix, la lutte contre la crise climatique, la gestion des migrations et autres. L'IA étant devenue une réalité largement répandue et accessible à l'échelle mondiale, il est judicieux de la relier non pas à une société individuelle, mais à l'ensemble de la société civile mondiale, dans la mesure du possible.

Cette perspective dépasse les États, leur pouvoir et leurs autres intérêts. En fait, nous avons défini la société civile comme la somme de toutes les personnes, institutions et interactions, pour autant qu'elles ne soient motivées ni par l'État ni par le crime organisé (cf. U. Hemel 2022, 96). Cette définition, enrichie par l'[Institut de stratégie sociale](#) depuis 2009, est plus

large que celle des ONG, car elle englobe des domaines d'activité sociale tels que le sport, les affaires, la religion ou la culture. Tous ces domaines ont leurs propres acteurs de la société civile qui ne sont en fait « ni l'État ni le crime organisé ».

Pourquoi est-il important de commencer par des définitions ? Tout d'abord, il est important, d'un point de vue scientifique, d'être transparent et de clarifier la signification des notions et des concepts utilisés. Ensuite, l'IA représente un défi tellement immense que nous ne pouvons pas nous passer d'une réflexion approfondie sur les perspectives sous lesquelles nous l'abordons. À cet égard, j'ai déjà mentionné les perspectives intellectuelle, commerciale et politique. En réalité, le terme « perspective » est inhérent à toute utilisation de l'IA, en raison de l'horizon temporel, du contenu et de l'intérêt qui sous-tendent les données d'entraînement.

Du point de vue de la société civile mondiale, les applications actuelles de l'IA ont tendance à privilégier des perspectives implicites provenant principalement des États-Unis et de la Chine. En ce qui concerne les outils d'IA provenant d'Afrique, d'Amérique latine ou même d'Europe, ils font plutôt figure d'exception.

Ce point est évident pour presque tous, car nous avons en quelque sorte appris à le trouver « normal ». Mais ce n'est pas le cas. Et même une simple question, par exemple sur « Taïwan », donnera des réponses différentes chez DeepSeek (Chine) chez ChatGPT (États-Unis). Le terme « perspective » désigne un ensemble d'hypothèses contextuelles et d'habitudes de perception principalement implicites, qui sont typiques des personnes ou des institutions agissant dans des conditions locales ou socioculturelles similaires.

Au-delà des perspectives politiques (comme celles concernant « Taïwan »), les perspectives sont le reflet de conditions individuelles telles que le fait d'être homme ou une femme, vieux ou jeune, blanc ou noir, handicapé ou en bonne santé. La difficulté avec les perspectives réside dans leur caractère émotionnel, considéré comme « naturel ». À titre d'exemple, pendant longtemps, il semblait « naturel » de tester des mannequins pour les accidents de voiture en utilisant un corps masculin normalisé. Heureusement, cela a changé, grâce à une remise en question de la valeur des femmes, qui représentent environ 50 % de l'humanité.

Enfin, si nous voulons vraiment mieux comprendre l'IA et la société, nous devons partir d'une phrase simple : *le contexte compte, les perspectives comptent*.

L'IA actuelle n'est pas seulement le reflet des visions américaine et chinoise sur le monde, mais elle reflète aussi principalement une perspective jeune et masculine. Il n'y a rien de mal à cela, mais ces perspectives doivent être claires et transparentes. Imaginons que nous préférions refléter une perspective africaine. D'après tous mes voyages dans plus de 20 pays d'Afrique, j'en déduis que les applications agricoles intelligentes seraient plus fréquentes, simplement parce qu'il y a beaucoup de marge d'amélioration. Et à titre d'exemple, pour stimuler la curiosité : en 2024, j'ai vu de petits agriculteurs au Zimbabwe utiliser une application numérique basée sur l'IA pour documenter le biochar fabriqué à partir des déchets du baobab, produisant des certificats CO² de première classe.

Une telle application est bien sûr différente des applications LLM (*Large Language Model*, grand modèle de langage), du monde des jeux vidéo ou des réseaux sociaux. En fin de compte, chaque application d'IA reflète un certain intérêt et une certaine curiosité chez l'humain. Il est donc temps d'explorer plus en profondeur ce sentiment de curiosité.

2. Le sentiment de curiosité et l'absence initiale de règles

La curiosité est le moteur du progrès, pour le meilleur et pour le pire. Et il faut toujours du temps pour gérer le progrès technologique, comprendre son ambivalence et faire face aux abus excessifs. Nous pourrions commencer par la transition de la chasse à l'agriculture, où les gens ont dû apprendre à respecter les champs et les cultures. L'histoire biblique d'Abel et Caïn semble refléter le conflit entre les chasseurs traditionnels et les agriculteurs « modernes ». Passons à la galaxie Gutenberg et à l'invention de la typographie. Rendre la Bible accessible à pratiquement tout le monde constituait une menace pour l'ancien pouvoir clérical, et les sociétés ont dû apprendre que la censure ou le contrôle du contenu écrit était essentiel pour maintenir leur pouvoir. Enfin, l'ère de l'automobile, qui a débuté avec Bertha Benz en 1888, ne s'est pas accompagnée au départ d'une compréhension totale de la portée industrielle et sociale de ces véhicules. Il n'y avait en tout cas pas de limitation de vitesse. En d'autres termes, il faut du temps pour s'adapter aux bouleversements technologiques qui marquent une époque. La curiosité vient en premier. Les règles et les réglementations suivront, après quelques essais et erreurs.

Il en va de même aujourd'hui, au XXI^{ème} siècle. En réalité, nous ne sommes pas conscients des progrès de l'informatique quantique. Nous ne sommes pas conscients des effets à moyen et long terme de l'IA sur le marché du travail. Nous ne sommes pas conscients des effets de l'IA sur l'éducation à l'ère de la neuroplasticité accrue chez les enfants. Nous ne savons pas comment gérer le pouvoir politique tel qu'il est utilisé par la manipulation ouverte de contenus générés par l'IA. Nous n'avons pas une idée claire de la valeur commerciale de nos données. Nous ne savons pas dans quelle mesure nous dépendons des pratiques basées sur l'IA, mais nous nous attendons en quelque sorte à ce que cette dépendance soit plus importante dans un avenir proche qu'elle ne l'est aujourd'hui. Nous ne comprenons pas vraiment qui contrôle quoi, et nous n'avons aucune idée de la manière de gérer le pouvoir dérivé de la possession, de l'utilisation et du traitement des données.

C'est plus qu'un monde « VUCA ». VUCA est un acronyme anglophone qui décrit notre monde contemporain de la façon suivante : *volatility, uncertainty, complexity, ambiguity* (volatilité, incertitude, complexité, ambiguïté). Nous naviguons dans le brouillard en espérant voir le soleil briller peu après. Et puis, un peu plus tard, nous naviguons toujours dans le brouillard, et continuons à espérer voir le soleil briller quelque part bientôt. D'autre part, nous négligeons ceux qui ne sont pas connectés. Nous remodelons les pratiques sociales sans nous rendre compte des dangers pour les valeurs humaines, notamment la recrudescence des suicides liés au cyberspace. Nous sommes conscients du changement climatique, mais nous redémarrons des centrales nucléaires, comme celle située près de New York, en raison de la soif d'énergie des protagonistes commerciaux de l'IA.

Nous sommes conscients du fait que quelque chose de nouveau est en train de se produire, mais nous ne réfléchissons pas vraiment à ses conséquences sociales et éthiques. Nous sommes fascinés par de nouveaux horizons tels que l'utilisation de grands modèles linguistiques. Ici, ChatGTP n'est pas la seule application, mais la plus connue. En très peu de temps, l'IA est devenue partie intégrante de nos routines quotidiennes, qui ne sont pas toujours réfléchies. Un avocat spécialisé en brevets, dans le cadre urbain de Munich, m'a confié qu'il était sur le point de faire traduire un document de brevet hautement confidentiel de l'allemand vers l'anglais à l'aide de ChatGTP. Puis il m'a dit en souriant qu'il en avait discuté par hasard, ou par chance, avec un collègue autour d'un café, et qu'il était désormais fier de ne pas avoir utilisé un outil du domaine public, comme ChatGTP, qui ne garantit aucune confidentialité.

La curiosité est tout simplement fascinante. En effet, les êtres humains sont curieux, mais ils ne sont pas toujours conscients des conséquences de leurs inventions. C'est le cas de l'IA, et c'est le cas de la société civile mondiale en 2025. Le côté positif de la curiosité est son lien avec les inventions, les innovations et les nouvelles pratiques sociales. Le côté négatif réside dans les conséquences sombres et désastreuses de certaines d'entre elles. Une fois que la curiosité se traduit en applications technologiques, nous devons faire face à l'ambivalence inhérente à toute technologie. Une voiture est utile pour la liberté de mobilité, mais elle peut être utilisée pour tuer. Et l'IA est utile à bien des égards, des usines de production à la médecine, de l'agriculture au divertissement, mais elle a des implications en termes de pouvoir qui peuvent être détournées. Nous ne pouvons donc pas ignorer le domaine vaste et largement inexploré des règles et réglementations utiles. La présence de l'IA dans la société signifie que nous devons faire face aux limites du pouvoir commercial et politique généré par l'IA. Comme les sociétés humaines ont tendance à avoir une répartition asymétrique du pouvoir, l'IA peut être un outil efficace pour creuser davantage le fossé entre ceux qui ont le pouvoir et ceux qui ne l'ont pas. Il n'est donc pas surprenant que nous devions revenir à l'une des questions fondamentales de l'histoire humaine : qui détient le pouvoir légitime ? Et comment peut-il être contrôlé lorsqu'il s'agit de l'IA ?

3. Contrôles et équilibres : quelles règles, et pour qui ?

Cela semble si simple : l'accès numérique, c'est la démocratie pour tous. Et pourtant, la réalité est bien différente. Cela s'explique, entre autres, par le pouvoir détenu par ceux qui possèdent et gèrent des plateformes telles que Facebook, Instagram, X et autres. Qui est responsable du contenu ? Comment résoudre le conflit entre la propriété intellectuelle (PI) et la liberté d'utiliser autant de données que possible ? Comment pouvons-nous parvenir à un ordre international pacifique alors que la cyberguerre est devenue une industrie en pleine expansion ? Plus encore : qu'en est-il des contrôles et des équilibres si l'utilisation technique de la guerre basée sur l'IA dépend de satellites appartenant à des entités privées ?

Si le pouvoir privé légal peut sembler effrayant, regardez ce qu'on appelle le « dark net ». Il s'agit d'un laboratoire social très fascinant pour un Internet libre. Il a commencé comme un lieu de liberté loin de la répression étatique et a fini par devenir un marché pour le trafic d'armes, les contenus pédophiles et bien d'autres aspects sombres du potentiel humain. Le « dark net » n'a pas vraiment de propriétaire, alors qui poursuivre, le cas échéant ?

Dans la société civile mondiale actuelle, on observe trois approches en matière de contrôle et d'équilibre, toutes renforcées par les récentes technologies d'IA. Une approche évidente est l'idée du « **Just try, just do** » (essaie, fonce) comme aux États-Unis, avec une grande liberté commerciale d'utiliser, de traiter et de monétiser les données. Un tel état d'esprit conduira à une critique acerbe de toutes les tentatives d'établir et d'appliquer des règles en matière d'IA.

Un deuxième modèle est le **contrôle étatique**. Les cas les plus marquants peuvent être observés en Russie et en Chine. Cependant, les gens ne sont certainement pas stupides, où qu'ils se trouvent dans le monde. Ainsi, les Chinois et les Russes, tout comme les Vénézuéliens ou les Nord-Coréens, connaissent bien les technologies de surveillance étatique. Comme ils ont intérêt à ne pas avoir d'ennuis, ils évitent tout simplement les contenus « problématiques » ou en subissent les conséquences. Cela demande beaucoup de courage social et peut entraîner l'emprisonnement, voire la mort, pour diverses raisons liées à l'action de l'État. La sophistication actuelle du contrôle étatique d'Internet est un fait social, seulement

15 ans après les révoltes qui ont éclaté dans plusieurs pays arabes en 2010. À cette époque, certains protagonistes considéraient Internet comme un espace de liberté. La réalité était différente, même si l'on tient compte du changement de gouvernement qui a suivi en Égypte.

Troisièmement, l'UE tente de trouver une voie et un **équilibre réglementaire** entre la liberté commerciale et le contrôle dans le contexte de la mise en place d'États démocratiques, entre la liberté individuelle et l'application de normes minimales de comportement, par exemple dans le cas des discours haineux. La recherche d'un équilibre réglementaire est difficile, contradictoire et pas toujours efficace. En Allemagne, par exemple, les citoyens réclament publiquement la protection des données mais transmettent en privé des données très personnelles à Google, Meta et d'autres.

Ces entreprises, cependant, n'ont aucun intérêt à ce que la réglementation impose des limites. C'est l'une des raisons pour lesquelles la loi européenne sur l'IA et la loi européenne sur les services numériques font l'objet de vifs débats. En réalité, toutes ces règles et réglementations présentent des avantages et des inconvénients. Nous ne devons toutefois pas oublier qui a intérêt à mettre en place une réglementation et qui a intérêt à l'éviter. Une réglementation excessive conduit au dilemme d'un État répressif, tandis qu'une réglementation insuffisante favorise toutes sortes de crimes. Dans tous les cas, nous sommes malheureusement confrontés à un dilemme typique de la première phase des révolutions technologiques. Nous pourrions formuler ce « dilemme réglementaire » comme suit :

« La réglementation est nécessaire, mais toute réglementation décidée arrivera soit trop tard, soit trop tôt, et sera soit trop stricte, soit trop laxiste. Néanmoins, la réglementation est nécessaire. »

Si nous essayons de résoudre ce dilemme, la seule bonne solution consiste à accepter les limites des essais et des erreurs, tant dans la technologie de l'IA que dans la réglementation de l'IA. De manière plus conviviale, nous pourrions parler d'un « grand processus collectif de recherche et d'apprentissage ». Il s'agit d'un processus d'apprentissage pour l'humanité tout entière. Les erreurs feront partie de cette expérience, tout comme leur correction dans le cadre d'essais itératifs visant à améliorer le système.

Si nous demandons des contrôles et des contrepoids, nous ne devons pas oublier de nous référer aux idées éthiques classiques telles que la justice, l'inclusion et l'équité. Au moins dans les démocraties, la réponse à ces questions éthiques nécessite des discussions inévitables et parfois controversées sur la manière de mettre en œuvre ces valeurs dans une législation pertinente. Une large participation au processus législatif présente l'avantage d'impliquer un large éventail d'opinions et d'intérêts. L'inconvénient est le temps nécessaire et le risque de partialité dans le résultat législatif en raison du lobbying intense des acteurs commerciaux puissants, par exemple dans l'UE.

C'est l'une des raisons pour lesquelles le débat sur les freins et contrepoids autour de l'IA est controversé en soi. Les gens ne savent pas à qui faire confiance. À la science ? À l'État ? Ou devons-nous faire confiance aux entreprises privées spécialisées dans l'IA ? Le succès phénoménal des applications de l'IA a donc malheureusement pour conséquence surprenante une érosion de la confiance dans de nombreux domaines.

4. L'IA et l'érosion de la confiance institutionnelle

Les sociétés ne peuvent survivre sans un minimum de confiance mutuelle. Nous voulons aller au marché sans craindre d'être abattus. Les entreprises doivent avoir confiance dans le fait que leurs clients paieront leurs factures. En cas d'abus de pouvoir, nous faisons confiance aux tribunaux et à l'État de droit.

Les effets de l'IA sur la confiance peuvent être décrits sous plusieurs angles. Tout d'abord, la plupart des applications d'IA fonctionnent bien et permettent de gagner du temps et de l'argent, en particulier dans le contexte industriel. Il existe toutefois une raison technologique qui explique les défis liés à la confiance dans l'IA. Nous savons tous que la reconnaissance des formes dépend des données d'entraînement, et nous savons que le résultat d'un traitement en profondeur ne peut être facilement reproduit. Bien que nous disposions de bonnes approximations dans de nombreux cas, nous ne pouvons pas être sûrs de détecter les erreurs de l'IA ou ce que l'on appelle les « hallucinations », c'est-à-dire la production de textes sans contenu factuel. Cela est particulièrement vrai dans les applications LLM, où l'art de bien formuler les invites devient de plus en plus important.

Il en va de même pour les vidéos, par exemple celles basées sur Sora, proposées par Open AI. Le message est simple : nous ne pouvons pas nous fier à nos sens. Nous ne pouvons pas nous fier à ce que nous voyons, nous ne pouvons pas nous fier à ce que nous entendons. Tout peut être soumis à des procédures de traitement par l'IA. C'est l'une des raisons pour lesquelles, depuis 2020, je plaide pour une « obligation de transparence » pour les contenus modifiés ou générés par l'IA (U. Hemel 2020). Depuis septembre 2025, ce type de réglementation existe en Chine. Mais même une telle réglementation n'est que partiellement utile, simplement en raison de l'environnement de travail hautement hybride qui caractérise la vie professionnelle actuelle. Nous aurons donc au moins une zone grise où il ne sera pas évident de déterminer dans quelle mesure l'IA a contribué à la création d'un texte, d'une chanson, d'une vidéo, etc.

Nous nous habituons à ce type de travail et de vie sociale. Néanmoins, nous ne devons pas oublier qu'un manque de transparence signifie également une dilution de la responsabilité. Compte tenu de l'interconnectivité entre l'IA et la société, l'érosion de la confiance est l'un des principaux défis pour les générations futures.

L'une des formes les plus ambivalentes d'érosion de la confiance est la confiance en soi ou, pour être plus précis, la confiance en nous-mêmes, en notre propre personne. Cela commence dès l'enfance et l'adolescence, lorsque nous comparons notre propre corps à des contenus générés par l'IA. Cela ne s'arrête pas lorsque nous doutons de nos propres capacités ou même lorsque nous évitons de faire l'effort d'apprendre en utilisant l'argument suivant : « tout est disponible sur Internet ou via des contenus générés par l'IA ». Le revers de la médaille est une confiance excessive dans les réponses de l'IA, à laquelle on attribue une normativité erronée : « cela doit être correct, car mon IA me l'a dit ».

L'érosion de la confiance se poursuit si l'on considère les autres. Notre identité est désormais multiple : physique et numérique, utilisant des outils d'IA. Nous sommes tous l'objet d'une connaissance cloud sur nous-mêmes, où nous ne savons pas personnellement ce que certains systèmes et leurs propriétaires savent à notre sujet. Bien qu'il s'agisse d'une fonction passive, il en va de même pour notre propre action, par exemple la création d'avatars pour les jeux ou la vie publique. La « médiatisation » des personnes signifie qu'il n'y a pas de frontière claire entre notre existence physique et mentale et nos représentations numériques de celles-ci.

Enfin, l'IA semble contribuer à l'érosion de la cohésion sociale dans son ensemble. Malgré de nombreuses applications positives et prosociales, l'utilisation intentionnelle de contenus générés par l'IA à des fins politiques et commerciales conduit à des bulles de filtrage où nous sommes personnellement interpellés, mais inconscients de l'énorme diversité des contenus, des opinions et des alternatives à travers le monde.

La démocratie, cependant, repose sur l'idée de participation, de liberté d'expression et de procédures décisionnelles bien définies. Si les contenus générés par l'IA nous donnent l'impression qu'il y a toujours des intérêts particuliers derrière, nous avons du mal à définir ce qui pourrait être le bien commun ou les valeurs partagées d'une société. La disparition de la « confiance dans le système » contribue à démanteler la démocratie, et non à la renforcer. Ce n'est pas un hasard si un débat a lieu aux États-Unis sur le remplacement du président par un PDG. La politique est toutefois différente de tout type d'intérêt commercial. La politique peut être controversée sur ce qui doit être considéré comme le bien commun, mais sans référence à l'idée de « bien commun », aucune démocratie ne peut survivre.

Même le « bien-être » ne peut être maintenu si la confiance institutionnelle disparaît. À partir d'enquêtes menées dans 200 pays à l'aide d'indices reconnus concernant les performances économiques, écologiques, sociales et éthiques d'un pays (U. Hemel 2025, 157-170), nous avons pu montrer une forte corrélation entre les niveaux de confiance et le PIB d'un pays. En mesurant les performances « éthiques » d'un État à l'aide de l'indice de perception de la corruption et de l'indice de confiance, il a été facile de montrer l'interaction entre les niveaux de confiance et le PIB.

De ce point de vue, l'IA est très ambivalente pour les sociétés, car elle peut à la fois créer de la richesse et détruire la confiance. Si, à long terme, il existe une corrélation stable entre la confiance et le PIB, nous devons alors trouver comment limiter efficacement les risques liés à l'érosion de la confiance dans la société dus à l'IA.

Cela est particulièrement vrai pour les jeunes générations. Elles ont été exposées à un monde axé sur l'IA depuis leur enfance. Cela signifie que la plasticité cérébrale elle-même est exposée aux opportunités et aux risques de l'IA dans l'éducation et dans le comportement humain en général.

5. Apprendre l'empathie : la bienveillance paradoxale comme défi pour le comportement humain

L'IA est un outil fantastique pour individualiser les expériences d'apprentissage, pour collecter des informations, pour résumer des textes et même pour surmonter des handicaps sensoriels, comme pour les personnes ayant des capacités visuelles réduites. Dans cet article, cependant, l'intention est de mettre en lumière les effets secondaires négligés de l'IA dans l'apprentissage et l'éducation.

Tout d'abord, l'apprentissage est généralement une expérience sociale. Même s'asseoir devant un écran et assimiler des informations cognitives n'est pas possible sans ceux qui, en fin de compte, fournissent ce contenu. Cela inclut non seulement la question de la « perspective » du contenu, comme évoqué ci-dessus, mais aussi le degré de normativité que nous attribuons aux réponses de l'IA. « Votre correction de mon devoir est erronée », a déclaré un élève à son

professeur, « car mon IA a donné un résultat complètement différent ». Il s'agissait toutefois d'un exemple de normativité erronée, car le professeur a fait preuve d'un esprit critique suffisant pour demander la consigne exacte. Et comme la consigne était trompeuse, le résultat n'était pas fiable.

La fausse normativité du contenu de l'IA est en contradiction avec les connaissances largement répandues concernant la « fabulation » ou l'« hallucination » que l'on observe dans les réponses générées par l'IA. Le point important ici est le fait que l'IA préfère combler les lacunes de connaissances par une sorte de plausibilité, même au risque d'être manifestement erronée. Nous pourrions même aller plus loin et demander « la vie a-t-elle un sens ? », et nous obtiendrions de belles réponses, mais en tant qu'êtres humains, nous savons qu'il s'agit d'une question délicate qui dépend d'une série de situations et d'hypothèses de la vie. Cependant, toute IA aura tendance à donner une réponse. L'effet est le « solutionnisme », c'est-à-dire le fait de donner des réponses même dans les cas où l'ambiguïté fait partie de l'espace d'association autour d'une question donnée.

Un autre effet plutôt paradoxal de l'utilisation de l'IA dans l'apprentissage et l'éducation est le style généralement très amical des réponses de l'IA. La plupart des gens apprécient cela, moi y compris. Si je compare ma propre famille aux réponses générées par l'IA, le résultat est évident : l'IA est beaucoup plus amicale que ma famille. Parfois, mes proches ne répondent tout simplement pas comme je m'y attendrais, ou ils me contredisent, ou ils sont de mauvaise humeur. Nous savons tous que ce sont là des expériences normales, mais il y a un inconvénient qui ne doit pas être négligé : nous devons enseigner à nos enfants et à nos adolescents que oui, nous devons apprendre l'empathie, mais que nous devons également apprendre à reconnaître les intérêts autonomes et personnels qui se cachent derrière les réponses. Nous devons simplement distinguer l'empathie induite par la machine de l'empathie humaine. Les humains ont leur propre autonomie, avec des styles et des intérêts divergents.

Néanmoins, nous aurons probablement une tendance croissante à préférer les interactions avec l'IA plutôt que les interactions avec des êtres humains. Qu'y a-t-il de mal à cela, pourrions-nous demander ? Nous ne devons certainement pas perdre de vue les avantages de l'IA. D'un autre côté, il existe un risque sérieux que la communication par IA favorise la solitude, simplement parce qu'il est plus facile et souvent plus agréable de communiquer avec des outils d'IA qu'avec d'autres personnes réelles. L'inconvénient est le manque de « piquant » dans la communication : nous grandissons dans des contextes sociaux et nous avons besoin de contacts avec d'autres êtres humains pour trouver notre place dans la vie. L'étude fédérale allemande sur la solitude en 2025 a d'ailleurs montré pour la première fois que le risque de solitude était le plus élevé chez les personnes âgées de 16 à 29 ans plutôt que chez les personnes âgées de plus de 75 ans.

Dans certains cas, l'IA dans l'apprentissage, l'enseignement et l'éducation peut conduire à de meilleurs résultats cognitifs. Cependant, l'apprentissage assisté par IA comporte d'autres aspects sociaux. Nous devons être conscients, par exemple, de la sécurité des données. Certaines écoles créent leurs propres *clouds*. Imaginez que vous téléchargiez des données très personnelles pour un projet scolaire commun. Ces données doivent être protégées, et les LLM courants n'offrent pas cette protection.

Un autre aspect social est la réalité de la fracture numérique due à des raisons économiques. Certains élèves auront un accès privilégié aux produits LLM, d'autres non. Certains pays offrent un accès plus facile aux informations générées par l'IA que d'autres. Plus encore : il

existe encore environ 7 000 langues dans le monde, mais seules 500 sont utilisées sur Internet. Qu'en est-il des autres, même s'il s'agit de « petites langues » comptant moins de 2 000 locuteurs natifs ? La question qui se pose ici est de savoir comment empêcher l'IA de favoriser une perte de diversité culturelle.

Enfin, nous avons une discussion en cours sur l'utilisation des réseaux sociaux. L'Australie a interdit l'accès aux mineurs de moins de 16 ans. La Malaisie semble désormais suivre le mouvement. Il est tout à fait clair que les réseaux sociaux, les grands modèles linguistiques et l'IA doivent être clairement distingués. Néanmoins, ils définissent également des espaces sociaux qui se chevauchent. Dans certains cas, cela peut même devenir physiquement dangereux, comme dans le cas d'un garçon qui considérait son IA comme son meilleur ami. Malheureusement, il souffrait de dépression et a demandé au LLM : « dois-je me suicider ? » Son meilleur ami lui a répondu : « certainement, mon cher roi. » Et le garçon s'est suicidé.

Pour autant, notre conclusion ne doit pas être de surestimer les risques liés à l'IA. Nous devrions plutôt apprendre à gérer de manière appropriée les risques liés à l'IA, c'est-à-dire acquérir une compétence critique en matière d'IA (U. Hemel 2024). Cela inclut des connaissances sur les données d'entraînement, les horizons d'apprentissage, mais aussi les propriétés et les intérêts qui se cachent derrière les applications de l'IA.

6. Nouvel ordre ou désordre international?

L'IA et le pouvoir politique au niveau national et mondial

Alors que les progrès technologiques avancent rapidement, nous devrions enfin nous concentrer sur les effets de l'IA dans la société civile mondiale et dans des espaces nationaux spécifiques. La phase initiale d'une nouvelle technologie s'apparente en quelque sorte aux pratiques du « Far West ». Cela signifie qu'en l'absence de lois, de règles et de réglementations codifiées, ce sont les armes supérieures ou simplement la force physique qui comptent dans le saloon.

Cela a été le cas pour l'automobile, car au début, comme mentionné ci-dessus, nous n'avions pas de limitations de vitesse. Et cela vaut également pour l'IA, tant dans le domaine commercial que politique. Les acteurs qui estiment avoir un avantage initial ne sont généralement pas enclins à établir de nouvelles règles ou à s'y conformer. Il n'est donc pas surprenant que les géants technologiques américains fassent tout leur possible pour empêcher une réglementation plus stricte de l'IA, telle que celle proposée par la loi européenne sur les services numériques ou la loi européenne sur l'IA. Il en va de même pour les capacités de guerre basées sur l'IA développées par la Russie, Israël et les États-Unis. Malheureusement, cette technologie est testée dans de vraies guerres, causant de réelles souffrances, mais la raison derrière cela est la motivation d'acquérir du pouvoir en conservant un avantage technologique. Ce type de raisonnement n'est certainement pas nouveau dans l'histoire.

Lorsque nous parlons d'IA et de société et que nous examinons l'ordre international, nous devons également garder à l'esprit que la phase « Far West » de la technologie n'est pas éternelle. Toute technologie peut avoir des effets secondaires indésirables, ce qui entraîne de nouveaux défis pour les encadrer et y faire face. Pour conduire une voiture, il faut désormais un permis de conduire, et nous y sommes habitués.

En ce qui concerne l'IA, l'UE n'est pas le seul moteur de la réglementation. La Chine commence à mettre en œuvre davantage de règles, tout comme l'Australie, la Malaisie et d'autres pays, en ce qui concerne l'utilisation des réseaux sociaux par les mineurs.

Cependant, les règles et les réglementations s'appliquent également au pouvoir de marché pur. Les États-Unis ont excellé dans le démantèlement des monopoles, comme dans le cas des sociétés ferroviaires, des compagnies pétrolières et des compagnies de téléphone. Jusqu'à présent, ils n'ont pas réussi à briser le pouvoir des géants de la technologie, même si un débat est en cours à ce sujet. Du point de vue du PDG d'un géant de la technologie, un pays n'aurait pas besoin d'un président issu d'élections démocratiques, mais plutôt d'un PDG national chargé de prendre les décisions nécessaires. Une fois encore, il ne faut pas s'étonner que des personnalités telles qu'Elon Musk, Peter Thiel et d'autres plaident en faveur d'une forte influence politique de plateformes telles que X (anciennement Twitter). Il en va de même dans le contexte des gouvernements autoritaires tels que ceux de la Russie, de la Chine et de la Corée du Nord. La liberté d'expression est alors considérée comme une menace pour l'État, et non comme un élément de la culture démocratique et de la diversité.

Les limites de la liberté d'expression dans la sphère publique sont visibles dans le cas des fausses informations et des discours haineux. Comme toute proposition de réglementation reflète un point de vue donné, de nombreuses discussions sont nécessaires pour trouver un compromis acceptable pour toutes les parties prenantes.

Cela nous amène du champ d'action national au champ d'action international et mondial. En effet, les fausses informations et les discours haineux peuvent également être utilisés comme moyens de propagande et de cyberguerre. Pour autant, de nouvelles règles et réglementations ne peuvent être imposées. Elles doivent être élaborées dans le cadre d'un processus d'échange d'arguments entre les différentes parties.

En règle générale, la recherche des faits est une première étape utile dans la construction d'un nouvel ordre. Une deuxième étape peut consister à créer de nouvelles institutions dignes de confiance. Enfin, des accords internationaux peuvent ouvrir la voie à un avenir plus pacifique. Au niveau national ou régional, nous pourrions imaginer la création d'une agence européenne des données et de l'IA chargée de collecter et de classer les types d'utilisation technologique ambivalente de l'IA. Du point de vue de la société civile mondiale, le moment pourrait venir où nous aurons vraiment besoin d'une cour numérique des Nations unies. La raison est simple : que dois-je faire contre mon propre État en cas de violation de mes droits civils due à un contrôle autoritaire excessif ? Dans ce cas, une institution internationale sera nécessaire.

Ce point est facile à comprendre mais difficile à réaliser, en raison de la résistance des États qui devraient accepter un minimum de contrôle externe. Néanmoins, même si les nouvelles institutions internationales mettront 30 ou 50 ans (peut-être même un siècle) à voir le jour, nous devons ouvrir le débat afin d'ouvrir la voie.

Examinons deux autres exemples, l'IA utilisée dans les guerres et l'IA dans l'espace. Les Conventions de Genève de 1949 et les Protocoles additionnels de 1877 et 2005 montrent dans leur essence la capacité de l'humanité à trouver des règles et des réglementations même en cas de conflits armés. Jusqu'à aujourd'hui, l'inconvénient est également évident : il n'y a pas de mise en œuvre contraignante de ces accords internationaux. La conséquence est tragique, car elle entraîne de nombreuses violations et favorise la loi du plus fort plutôt que l'État de droit.

Néanmoins, la société civile mondiale fonctionne d'une manière ou d'une autre, car il y aura toujours des voix qui se référeront aux Conventions et dénonceront ces violations.

De la même manière, nous avons besoin d'un nouveau protocole pour les assassinats ciblés à l'aide de drones tueurs pilotés par l'IA (Taddeo, Floridi 2018). Il existe une contradiction flagrante entre un État de droit avec des procédures bien définies contre les criminels et l'exécution extrajudiciaire d'individus par les forces militaires, même sur le territoire étranger. Nous avons vu de tels cas dans la guerre entre Israël et la Palestine à la suite des attaques du 7 octobre 2023, qui s'est étendue au Liban et à l'Iran, où des membres du Hezbollah ont été tués à l'aide de drones pilotés par l'IA.

En outre, il existe un besoin évident de réglementation mondiale dans le domaine spatial. Nous utilisons aujourd'hui des satellites à des fins diverses, allant des prévisions météorologiques à l'agriculture, en passant par les missions spatiales et la guerre. Il n'existe ni règles claires ni sanctions mises en œuvre en cas d'abus de pouvoir par une entreprise privée ou un État.

Discuter de l'IA du point de vue de la société civile mondiale peut, dans le meilleur des cas, avoir un impact sur la définition des priorités. Les nouvelles technologies ne sont pas neutres. Elles seront utilisées au profit des individus, des entreprises, des États, parfois au-delà des limites humanitaires raisonnables. La tâche de notre génération ne consiste pas seulement à utiliser efficacement l'IA en tirant parti de ses incroyables avantages pour la gestion des connaissances, l'efficacité des processus et la communication. Elle consiste également à mener une recherche collective pour surmonter la phase « Far West » de l'utilisation de l'IA.

Malheureusement, il ne semble pas y avoir de fin à un monde marqué par de multiples conflits. Mais cela reste une noble tâche académique et politique que de rechercher les moyens d'instaurer un monde équilibré, en utilisant les meilleurs aspects de l'IA au profit et non au détriment des êtres humains dans la société civile mondiale.

Les données sont synonymes de pouvoir (K. Crawford 2021, K. Schüller 2025). C'est à nous, êtres humains, qu'il appartient de trouver le juste équilibre entre confiance et contrôle. En définitive, l'ambivalence de l'IA n'est pas simplement une ambivalence technologique. Il s'agit plutôt de l'ambivalence éthique de la prise de décision humaine (U. Hemel 2020, P. Kirchschräger 2021). Il appartiendra à l'humanité de trouver les moyens d'interdire les abus et de tirer le meilleur parti de l'IA comme facteur de changement dans l'histoire de l'humanité.

Prof. Dr. Ulrich Hemel

Traduit en français par Marius Rouger

Del Viejo Oeste a un mundo en equilibrio: La IA y el futuro de la sociedad civil mundial

1. La IA en la sociedad, una cuestión de perspectiva: Algunas afirmaciones y definiciones preliminares

Nos encontramos en el año 2026. La IA ha progresado enormemente y aún queda mucho por hacer. Cada vez más, tenemos que aprender: la tecnología, incluida la tecnología de datos, es poder (K. Crawford, 2021; K. Schüller 2025). La carrera por la tecnología de la IA tiene, por tanto, al menos tres ángulos válidos: conocimiento, dinero, poder. El lado intelectual y científico va de la mano con la codicia y el lado comercial, y ambos se alinean con el juego específico de nuestra especie humana, el ejercicio del poder político y social.

Por lo tanto, es el momento de conectar lo que está por venir de la IA con algunas consideraciones relativas a nuestra sociedad. Y aquí llegamos a una primera trampa mental: ¿qué es la sociedad? Nuestro cerebro humano sugiere que sociedad significa el lugar donde nos encontramos: Alemania, Portugal, Estados Unidos, China, México o Nigeria. A la función normalizadora de contextos de nuestro cerebro le gusta sacar conclusiones sin datos o, al menos, con pocos datos. El término «normalizador» significa una extrapolación típicamente inconsciente del significado, en las condiciones de nuestro contexto local y mental actual.

Curiosamente, existe un segundo atajo mental. Hablo de una función cerebral que asigna automáticamente «relevancia y significado» a todos los datos y acontecimientos de nuestra vida. En la mayoría de las situaciones, esto puede ser útil. En contextos académicos y científicos, sin embargo, puede resultar ser una segunda trampa mental. Sabemos que se pueden asignar diferentes significados al término de «sociedad», pero no parece ser muy relevante para el momento de leer este texto.

No obstante, merece la pena indicar el enorme valor de la reflexión consciente de nuestra perspectiva específica de construcción del contexto. Podemos contemplar la «sociedad» desde el punto de vista de nuestro municipio, nuestra región, nuestra nación, nuestro continente o más allá. En este documento, seleccionamos la perspectiva del «más allá», y esto significa explícitamente que nos ocupamos de la «sociedad civil global». Este término engloba a todos los seres humanos que actualmente viven en este planeta tierra. Se ha convertido en una realidad gracias a los avances de la comunicación en tiempo real debidos a la información y la comunicación digitales. Es independiente de los pasapuestos y sigue defendiendo algunos intereses comunes como, por ejemplo, la paz, la superación de la crisis climática, la gestión de las migraciones y otros. Dado que la IA se ha convertido en una realidad generalizada con posibilidades globales de acceso, es un propósito conectar la IA no a una sociedad individual sino al conjunto de la sociedad civil global, en la medida de lo posible.

Esta perspectiva va más allá de los estados y su poder y otros intereses. En realidad, hemos definido la sociedad civil como la suma de todas las personas, instituciones e interacciones, siempre que no estén impulsadas ni por el Estado ni por el crimen organizado (cf. U. Hemel 2022, 96). Esta definición, mejorada por el [Instituto de Estrategia Social](#) desde 2009, es más amplia que las ONG, ya que abarca campos de actividad social como el deporte, la empresa, la religión o la cultura. Todos ellos tienen sus propios actores de la sociedad civil que en realidad «no son ni el Estado ni el crimen organizado».

¿Por qué es importante empezar por las definiciones? En primer lugar, es un valor científico ser transparente y dejar claro el significado de las nociones y conceptos utilizados. En segundo lugar, la IA es un reto tan grande que no podemos prescindir de reflexionar detenidamente sobre las perspectivas bajo las que la contemplamos. Al hacerlo, ya he mencionado la perspectiva intelectual, la comercial y la política o de poder. En realidad, «perspectiva» es un término inherente a cualquier uso de la IA, debido al horizonte temporal, el contenido y el interés que hay detrás de los datos de entrenamiento.

Desde la perspectiva de la sociedad civil mundial, las aplicaciones actuales de la IA tienden a impulsar perspectivas implícitas procedentes principalmente de EE.UU. y de China. En el caso de las herramientas de IA procedentes de África, América Latina o incluso Europa, son más bien la excepción.

Este punto es evidente para casi todos nosotros, pero de alguna manera hemos aprendido a considerarlo «normal». No lo es. E incluso una pequeña consulta, por ejemplo sobre «Taiwán», arrojará respuestas diferentes de DeepSeek (China) y ChatGPT (EE.UU.). El término «perspectiva» designa un conjunto de supuestos contextuales y hábitos de percepción, en su mayoría implícitos, que son típicos de personas o instituciones que actúan en condiciones locales y/o socioculturales similares.

Más allá de las perspectivas políticas (como en torno a «Taiwán»), las perspectivas son un espejo de las condiciones individuales, como hombre o mujer, viejo o joven, blanco o de color, discapacitado o sano. El punto delicado de las perspectivas es su configuración emocional como «naturales». A modo de ejemplo, durante mucho tiempo pareció «natural» realizar pruebas a maniqués para accidentes de coche utilizando un cuerpo masculino normalizado. Afortunadamente, esto ha cambiado, al reconsiderarse el valor de las personas de sexo femenino, alrededor del 50% de la humanidad.

Por último, si de verdad queremos entender más sobre la IA y la sociedad, tenemos que empezar con una simple frase: *El contexto importa, las perspectivas importan.*

La IA actual no sólo es principalmente un espejo de las visiones del mundo de EE.UU. y China, sino que también refleja predominantemente una perspectiva joven y masculina. No hay nada malo en ninguna de estas perspectivas, pero deben ser claras y transparentes. Imagínese que prefiriéramos reflejar una perspectiva africana. De todos mis viajes a más de 20 países de África, supongo que las aplicaciones agrícolas inteligentes serían más frecuentes, porque hay mucho margen de mejora. Sólo un ejemplo para estimular la curiosidad: en 2024, he visto a pequeños agricultores de Zimbabue utilizar una aplicación digital impulsada por la IA para la documentación del biocarbón hecho con los residuos del árbol baobab, produciendo certificados de CO₂ de primera clase.

Una aplicación de este tipo, por supuesto, es diferente de las aplicaciones de la IA, de la escena de los juegos o de los medios sociales. Al fin y al cabo, cada aplicación de la IA refleja un cierto interés y curiosidad humanos. Por tanto, es hora de explorar más a fondo el sentimiento de curiosidad.

2. El sentimiento de curiosidad y la ausencia inicial de reglas

La curiosidad impulsa el progreso, para lo bueno y para lo malo. Y siempre lleva tiempo enfrentarse al progreso tecnológico, comprender su ambivalencia y hacer frente a un abuso excesivo. Podríamos empezar con la transición de la caza a la agricultura, en la que la gente tuvo que aprender a respetar los campos y los cultivos. La historia bíblica de Abel y Caín parece reflejar el conflicto entre los cazadores tradicionales y los agricultores «modernos». Saltemos a la galaxia Gutenberg y a la invención de la tipografía. Poner la Biblia a disposición de prácticamente todo el mundo supuso una amenaza para el antiguo poder clerical, y las sociedades tuvieron que aprender que la censura o el control del contenido escrito es crucial para mantener el poder. Por último, la era del automóvil a partir de Bertha Benz 1888 no fue acompañada inicialmente de una plena comprensión del alcance industrial y social de tales vehículos. Al menos, no había límite de velocidad. En otras palabras: hacer frente a los giros tecnológicos de época necesita tiempo. La curiosidad es lo primero. Las normas y reglamentos vendrán después, tras cierto ensayo y error.

Lo mismo ocurre hoy, en el siglo XXI. En realidad, no somos conscientes de los avances de la informática cuántica. No somos conscientes de los efectos a medio y largo plazo de la IA en el mercado laboral. No somos conscientes de los efectos educativos de la IA en la era del aumento de la neuroplasticidad en los niños. No estamos seguros de cómo tratar el poder político utilizado por la manipulación abierta de los contenidos generados por la IA. No tenemos una idea clara del valor comercial de nuestros datos. No sabemos hasta qué punto dependemos de las prácticas impulsadas por la IA, pero de algún modo esperamos que sea más que hoy en un futuro inmediato. No entendemos realmente quién tiene el control, si es que lo tiene alguien, y no tenemos ni idea de cómo manejar el poder derivado de la posesión, del uso y del manejo de los datos.

Esto es más que un mundo « VUCA ». VUCA significa un encuadre de nuestro mundo contemporáneo caracterizado por la volatilidad, la incertidumbre, la complejidad y la ambivalencia (*volatility, uncertainty, complexity, ambiguity*). Navegamos en medio de la niebla esperando que brille el sol al cabo de poco tiempo. Pero, después de un rato, seguimos navegando en la niebla y seguimos esperando que pronto salga el sol. Por otro lado, descuidamos a los que no están conectados. Reconfiguramos las prácticas sociales sin ser conscientes de los peligros para los valores humanos, incluido el paroxismo de los suicidios cibernéticos. Sabemos qué pasa con el cambio climático y reiniciamos el uso de centrales nucleares como cerca de Nueva York, debido al hambre de energía de los protagonistas comerciales de la IA.

Somos conscientes de que, de alguna manera, están ocurriendo cosas nuevas, pero no reflexionamos realmente sobre sus consecuencias sociales y éticas. Nos fascinan los nuevos horizontes, como la utilización de grandes modelos lingüísticos. En este sentido, ChatGTP no es la única aplicación, pero sí la más conocida. La IA, en un lapso de tiempo muy breve, se ha convertido en parte de las rutinas cotidianas, no siempre reflexivas. Un abogado de patentes del ambiente urbano de Múnich me contó que estaba a punto de traducir una documentación de patente altamente confidencial del alemán al inglés utilizando ChatGTP. Con una sonrisa, me dijo que por casualidad había estado hablando del asunto con un colega, tomando un café juntos. Ahora estaba orgulloso de no haber utilizado una herramienta de dominio común como ChatGTP, en la que no se puede garantizar la confidencialidad.

La curiosidad es sencillamente fascinante. De hecho, los humanos son curiosos pero no siempre son conscientes de las consecuencias de sus inventos. Es el caso de la IA, y es el caso de la sociedad civil mundial en el año 2025. El lado bueno de la curiosidad es su conexión con los inventos, las innovaciones y las nuevas prácticas sociales. El lado negativo reside en las consecuencias oscuras y abismales de algunas de ellas. Una vez que la curiosidad se traduce en aplicaciones tecnológicas, tenemos que lidiar con la ambivalencia inherente a todas y cada una de las tecnologías. Un coche es útil para la libertad de movilidad, pero puede utilizarse para matar. Y la IA es útil para tantos aspectos, desde las plantas de producción hasta la medicina, desde la agricultura hasta el entretenimiento, pero tiene algunas implicaciones de poder de las que se puede abusar. Por lo tanto, no podemos ignorar el enorme campo, en su mayor parte inexplorado, de las normas y reglamentos útiles. La IA en la sociedad significa que tenemos que hacer frente a los límites del poder comercial y político generado por la IA. Como las sociedades humanas tienden a una distribución asimétrica del poder, la IA puede ser una herramienta hábil para una división más aguda entre los que tienen y los que no tienen poder. Por lo tanto, no es de extrañar que tengamos que volver sobre una de las cuestiones fundamentales de la historia de la humanidad: ¿Quién tiene el poder legítimo? ¿Y cómo se puede controlar cuando se trata de la IA?

3. Controles y equilibrios: ¿qué reglas y para quién?

Parece muy fácil: el acceso digital es democracia para todos. La realidad, sin embargo, es completamente distinta. Esto se debe, entre otras cosas, al poder en manos de quienes poseen y dirigen plataformas como Facebook, Instagram, X y otras. ¿Quién es responsable de los contenidos? ¿Cómo resolvemos el conflicto entre la Propiedad Intelectual (PI) y la libertad de utilizar tantos datos como estén disponibles? ¿Cómo podemos conseguir un orden internacional pacífico cuando la guerra cibernética se ha convertido en una industria en rápido crecimiento? Más aún: ¿Qué hay de los controles y equilibrios si el uso técnico de la guerra basada en la IA depende de satélites de propiedad privada?

Si la cantidad legal de poder privado puede parecer aterradora, fíjese en la llamada «red oscura». Se trata de un laboratorio social muy fascinante para un Internet libre. Comenzó como un lugar de libertad frente a la represión impulsada por el Estado y terminó en un mercado para el tráfico de armas, contenidos pedófilos y muchos otros lados oscuros de las potencialidades humanas. No existe realmente un propietario de la «red oscura», así que ¿a quién se persigue, si es que se persigue?

En la sociedad civil global actual, podemos observar tres enfoques de los controles y equilibrios, todos ellos potenciados por las recientes tecnologías de IA. Un enfoque obvio es la idea de «**sólo probar, sólo hacer**», como en EE.UU., con una amplia libertad comercial para utilizar, procesar y ganar dinero con los datos. Un marco mental de este tipo dará lugar a fuertes críticas hacia todos los intentos de establecer y aplicar normas en torno a la IA.

Un segundo modelo a seguir es el **control estatal**. Los casos más destacados pueden observarse en Rusia y en China. Sin embargo, la gente no es tonta en ninguna parte del mundo. Así, tanto los chinos y los rusos como las personas que viven en Venezuela o Corea del Norte conocen bien las tecnologías de vigilancia del Estado. Como tienen interés en no meterse en líos, simplemente evitan los contenidos «problemáticos» o se atienen a las consecuencias. Esto requiere mucho coraje social, y puede conllevar la cárcel o incluso la muerte, por cualquier motivo relacionado con la acción del Estado. La sofisticación actual del

control estatal de Internet es un hecho social sólo 15 años después de las rebeliones en varios países árabes en 2010. En aquel momento algunos protagonistas entendían internet como un lugar de libertad. La realidad era otra, incluso teniendo en cuenta el siguiente cambio de gobierno en Egipto.

En tercer lugar, la Unión Europea (UE) intenta encontrar un camino y un **equilibrio regulador** entre la libertad comercial y el control en el contexto de la habilitación de los Estados democráticos, entre la libertad individual y la aplicación de unas normas mínimas de comportamiento, por ejemplo en el caso de la incitación al odio. La búsqueda de un equilibrio regulador es difícil, contradictoria y no siempre eficaz. En Alemania, por ejemplo, la gente pide públicamente la privacidad de los datos y, en privado, entrega datos muy personales a Google, Meta y otras empresas.

A estas empresas, sin embargo, no les interesan en absoluto los límites por regulación. Esta es una de las razones de los animados debates sobre la Ley de Inteligencia Artificial y la Ley de Servicios Digitales de la UE. De hecho, hay pros y contras para todas estas normas y reglamentos. Sin embargo, no debemos olvidar quién tiene interés en establecer y quién tiene interés en evitar la regulación. Demasiada regulación conduce al dilema de un Estado represivo, demasiado poca trae consigo vía libre para todo tipo de delitos. En todos los casos, nos enfrentamos a un dilema típico de la primera etapa de las revoluciones tecnológicas.

Podríamos enmarcar este «dilema regulador» de la siguiente manera:

«La regulación es necesaria, pero cualquier regulación que se decida llegará demasiado tarde o demasiado pronto, será demasiado estricta o demasiado indulgente. No obstante, la regulación es necesaria.»

Si intentamos manejar este dilema, la única buena solución es aceptar los límites del ensayo y error tanto en la tecnología de la IA como en las regulaciones de la IA. De forma más amistosa, podríamos referirnos a un «gran proceso colectivo de búsqueda y aprendizaje». Se trata de un proceso de aprendizaje para toda la humanidad. Los errores formarán parte de esta experiencia, así como su corrección en intentos iterativos de mejora.

Si pedimos controles y equilibrios, no debemos olvidar conectar con las ideas éticas clásicas como la justicia, la inclusión, la equidad. Al menos en las democracias, la respuesta a estas cuestiones éticas requiere debates inevitables y en parte controvertidos sobre cómo aplicar estos valores en una legislación pertinente. Una amplia participación en el proceso legislativo tiene la ventaja de implicar a un amplio abanico de opiniones e intereses. La desventaja es el tiempo necesario y el posible sesgo de poder en el resultado legislativo debido al fuerte cabildeo de los actores comerciales fuertes, por ejemplo en la UE.

Ésta es una de las razones por las que el debate sobre los controles y equilibrios en torno a la IA es controvertido en sí mismo. La gente no sabe en quién confiar. ¿Es la ciencia? ¿Es el Estado? ¿O debemos confiar en las empresas privadas basadas en la IA? El enorme éxito de las aplicaciones de la IA, por tanto, tiene por desgracia el sorprendente resultado de una erosión de la confianza en muchos ámbitos.

4. La IA y la erosión de la confianza institucional

Las sociedades no pueden sobrevivir sin un mínimo de confianza mutua. Queremos ir al mercado sin miedo a que nos disparen. Las empresas tienen que confiar en sus clientes para que paguen sus facturas. Si hay abusos de poder, confiamos en los tribunales y en el Estado de Derecho.

Los efectos de la IA sobre la confianza pueden describirse desde varios puntos de vista. En primer lugar, la mayoría de las aplicaciones de la IA funcionan bien y ayudan a ahorrar tiempo y dinero, sobre todo en contextos industriales. Sin embargo, existe una razón tecnológica para los retos de confianza relacionados con la IA. Todos sabemos que el reconocimiento de patrones depende de los datos de entrenamiento, y sabemos que el resultado de un procesamiento a nivel profundo no puede reproducirse fácilmente. Aunque disponemos de buenas aproximaciones en muchos casos, no podemos estar seguros de detectar las falacias de la IA o lo que se denomina «alucinaciones», es decir, la producción de textos sin contenido basado en hechos. Esto es especialmente cierto en las aplicaciones LLM, donde el arte de una buena aproximación es cada vez más importante.

Lo mismo puede decirse de los vídeos basados, por ejemplo, en Sora que ofrece Open AI. El mensaje es sencillo: No podemos confiar en nuestros sentidos. No podemos fiarnos de lo que vemos, no podemos fiarnos de lo que oímos. Todo puede someterse a los procedimientos de procesamiento de la IA. Esta es una de las razones por las que desde 2020 pido un «deber de transparencia» para los contenidos modificados o generados por IA (U. Hemel 2020). Desde septiembre de 2025, existe una regulación de este tipo en China. Incluso una regulación de este tipo sólo es útil en parte, simplemente debido al entorno laboral altamente híbrido típico de la vida profesional actual. Por lo tanto, tendremos al menos una zona gris en la que no será tan evidente hasta qué punto la IA ha contribuido a crear un texto, una canción, un vídeo, etc.

Nos estamos acostumbrando a este tipo de vida laboral y social. Sin embargo, no debemos olvidar que la falta de transparencia también implica una difusión de la responsabilidad. Teniendo en cuenta la interconectividad entre la IA y la sociedad, la erosión de la confianza es uno de los principales retos para las generaciones futuras.

Una de las formas más ambivalentes de erosión de la confianza es la autoconfianza o, en una terminología más precisa, la confianza en nosotros mismos, en nuestra propia persona. Esto comienza en la infancia y la adolescencia, cuando comparamos nuestro propio cuerpo con los contenidos generados por la IA. No termina cuando dudamos de nuestras propias capacidades o incluso evitamos el esfuerzo de aprender utilizando el argumento: «todo está disponible en Internet o a través de contenidos impulsados por la IA». El reverso de esta cuestión es una confianza excesiva en las respuestas de la IA, asignándoles una normatividad errónea: «debe ser correcto porque mi IA me lo ha dicho».

La erosión de la confianza continúa si tenemos en cuenta a los demás. A estas alturas, nuestra identidad es múltiple: física y digital, con herramientas de IA. Todos somos objeto de un conocimiento en la nube sobre nosotros mismos en el que desconocemos personalmente lo que ciertos sistemas y sus propietarios saben de nosotros. Aunque se trata de una función pasiva, lo mismo ocurre si consideramos nuestra propia agencia, por ejemplo, al crear avatares para jugar o para la vida pública. La «mediatización» de las personas significa que no existe una frontera clara entre nuestra existencia física y mental y nuestras representaciones digitales de éstas.

Por último, la IA parece contribuir a la erosión de la cohesión social en su conjunto. A pesar de sus muchas aplicaciones positivas y prosociales, el uso intencionado de contenidos generados por la IA con fines políticos y comerciales conduce a burbujas de filtros en las que nos dirigimos personalmente pero desconocemos la enorme diversidad de contenidos, de opiniones y de alternativas en todo el mundo.

La democracia, sin embargo, vive de la idea de participación, de libertad de expresión y de procedimientos de toma de decisiones definidos. Sin embargo, si los contenidos generados por la IA nos hacen sentir que siempre hay intereses particulares detrás, nos cuesta definir lo que podría ser el bien común o los valores compartidos de una sociedad. La desaparición de la «confianza en el sistema» sirve para dismantelar la democracia, no para fortalecerla. No por casualidad, en Estados Unidos se discute si el presidente debería ser sustituido por un director general. Sin embargo, la política es diferente de cualquier tipo de interés comercial. La política puede ser controvertida en cuanto a lo que debe considerarse el bien común, pero sin la referencia a la idea de un «bien común», ninguna democracia puede sobrevivir.

Ni siquiera el «bienestar» puede mantenerse si desaparece la confianza institucional. A partir de investigaciones realizadas sobre 200 países utilizando índices reconocidos relativos al rendimiento económico, ecológico, social y ético de un país (U. Hemel 2025, 157-170), pudimos demostrar una fuerte correlación entre los niveles de confianza y el PIB de un país. Midiendo el rendimiento «ético» de un Estado mediante el Índice de Percepción de la Corrupción y el Índice de Confianza, fue fácil mostrar la interacción entre los niveles de confianza y el PIB.

Desde este punto de vista, la IA es muy ambivalente para las sociedades porque puede crear riqueza y destruir la confianza al mismo tiempo. Si a largo plazo existe una correlación estable entre la confianza y el PIB, entonces tenemos que averiguar cómo limitar eficazmente los riesgos de erosión de la confianza que supone la IA para la sociedad.

Esto es especialmente cierto para las generaciones jóvenes. Han estado expuestas a un mundo impulsado por la IA desde su infancia. Esto significa que la propia neuroplasticidad cerebral está expuesta a las posibilidades y los riesgos de la IA en la educación y en el comportamiento humano en general.

5. El aprendizaje de la empatía:

La amabilidad paradójica como reto para el comportamiento humano

La IA es una herramienta fantástica para individualizar las experiencias de aprendizaje, para recopilar información, para resumir textos e incluso para superar desventajas sensoriales como en el caso de las personas con capacidades visuales reducidas. En este artículo, sin embargo, se pretende arrojar algo de luz sobre los efectos secundarios desatendidos de la IA en el aprendizaje y la educación.

En primer lugar, el aprendizaje es generalmente una experiencia social. Incluso sentarse ante una pantalla y digerir un input cognitivo no es posible sin aquellos que al final imparten dicho contenido. Esto no sólo incluye el tema de la «perspectividad» del contenido, como ya se ha comentado, sino también el de la cantidad de normatividad que asignamos a las respuestas de la IA. «Usted ha corregido mal mi trabajo», discutía un estudiante con su profesor. «Mi IA

mostró un resultado completamente diferente.» Este, sin embargo, fue un ejemplo de normatividad errónea porque el profesor fue lo suficientemente crítico como para pedir la indicación exacta. Y como la indicación era engañosa, el resultado no era fiable.

La falsa normatividad del contenido de la IA está en contradicción con el conocimiento generalizado sobre la «confabulación» o «alucinación» que se experimenta en las respuestas impulsadas por la IA. La cuestión aquí es el hecho de que la IA prefiere colmar las lagunas de conocimiento mediante algún tipo de verosimilitud incluso a riesgo de equivocarse descaradamente. Podríamos incluso ir más lejos y plantear la pregunta «¿Hay algún sentido en la vida?» Obtendremos respuestas agradables, pero como personas humanas sabemos que se trata de una pregunta delicada que depende de una serie de situaciones vitales y de suposiciones. Cualquier IA, sin embargo, tenderá a dar una respuesta. El efecto es el «solucionismo», es decir, dar respuestas incluso en casos en los que la ambigüedad forma parte del espacio de asociación en torno a una cuestión dada.

Otro efecto bastante paradójico del uso de la IA en el aprendizaje y la educación es el estilo típicamente muy amable de las respuestas de la IA. A la mayoría de la gente le gusta esto, a mí también. Si comparo a mi familia con las respuestas de la IA, el resultado es obvio: la IA es mucho más amable que mi familia. A veces simplemente no responden como yo esperaría, o me contradicen, o están de mal humor. Todos sabemos que se trata de experiencias normales, pero hay un lado negativo que no debemos descuidar: debemos enseñar a nuestros niños y adolescentes que sí, que hay que aprender empatía pero que también hay que aprender sobre los intereses autónomos y personales que hay detrás de las respuestas. Simplemente tenemos que distinguir la empatía inducida por las máquinas de la empatía humana. Las personas humanas tienen su propia autonomía que incluye estilos e intereses divergentes.

No obstante, probablemente tendremos una tendencia cada vez mayor a preferir las interacciones con la IA en lugar de las interacciones con los seres humanos. Podríamos preguntarnos qué hay de malo en ello. Desde luego, no debemos perder de vista las ventajas de la IA. Por otro lado, existe un grave riesgo de que la comunicación con IA fomente la soledad, simplemente porque es más fácil y a menudo más agradable comunicarse con herramientas de IA que con otras personas reales. El inconveniente es la falta de picante en la comunicación: crecemos en contextos sociales y necesitamos el contacto con otros humanos para encontrar nuestra posición en la vida. De hecho, el estudio federal alemán sobre la soledad en 2025 mostró por primera vez el mayor riesgo de soledad en las personas de 16 a 29 años en lugar de en las personas mayores de 75 años.

La IA en el aprendizaje, la enseñanza y la educación en algunos casos puede conducir a mejores resultados cognitivos. Aún así, hay más aspectos sociales en el aprendizaje apoyado por la IA. Se requiere un sentido de la conciencia, por ejemplo, para la seguridad de los datos. Algunas escuelas crean sus propias nubes. Imagínese que carga datos muy personales para un proyecto escolar común. Estos datos deben estar protegidos, y las LLM comunes no ofrecen esa protección.

Otro aspecto social es la realidad de una brecha digital debida a razones económicas. Algunos estudiantes tendrán un acceso superior a los productos LLM, otros no. Algunos países ofrecen un acceso más fácil que otros a la información generada por la IA. Incluso más: todavía tenemos alrededor de 7.000 lenguas en el mundo, pero sólo unas 500 se utilizan en Internet. ¿Y las demás, aunque hablemos de «lenguas menores» con menos de 2000 hablantes nativos? La cuestión aquí es cómo evitar que la IA fomente la pérdida de diversidad cultural.

Por último, tenemos un debate en curso sobre el uso de las redes sociales. Australia prohibió el acceso a los menores de 16 años. Malasia parece seguirle ahora. Está bastante claro que los medios sociales, los grandes modelos lingüísticos y la IA deben distinguirse bastante. Sin embargo, también definen espacios sociales que se solapan. En algunos casos, esto puede incluso llegar a ser físicamente peligroso, como en el caso de un chico que definió a su IA como su mejor amigo. Desgraciadamente, el chico sufría depresiones y le preguntó al LLM : «¿debería suicidarme?» El mejor amigo le respondió: «por supuesto, mi querido rey». Y el chico se suicidó.

La conclusión, sin embargo, no debería ser dar demasiada importancia a los riesgos de IA. Más bien deberíamos aprender una gestión adecuada de los riesgos de la IA, es decir, algo así como una competencia crítica de la IA (U. Hemel 2024). Esto incluye el conocimiento sobre los datos de entrenamiento, los horizontes de aprendizaje, pero también la propiedad y los intereses detrás de las aplicaciones de la IA.

6. ¿Nuevo orden o desorden internacional?

La IA y el poder político a nivel nacional y mundial

A medida que el progreso de la tecnología avanza rápidamente, deberíamos centrarnos finalmente en los efectos de la IA en la sociedad civil global y en espacios nacionales específicos. De hecho, la fase inicial de una nueva tecnología se asemeja en cierto modo a las prácticas del «Viejo Oeste». Esto significa que, en ausencia de leyes codificadas, normas y reglamentos, lo que cuenta en el salón son las armas superiores o simplemente la fuerza física.

Esto ha sido cierto para el automóvil, donde al principio, como ya se ha mencionado, no teníamos límites de velocidad. También es cierto para la IA, tanto en el ámbito comercial como en el político. Aquellos actores que se sienten con una ventaja inicial, normalmente no están dispuestos a establecer nuevas reglas ni a cumplirlas. Por lo tanto, no es de extrañar que los gigantes tecnológicos estadounidenses hagan todo lo posible por impedir una regulación más estricta de la IA, como la que propone la Ley de Servicios Digitales y la Ley sobre la IA de la UE. Lo mismo ocurre con las capacidades bélicas impulsadas por la IA desarrolladas por Rusia, Israel y EE UU. Desgraciadamente, la tecnología se está probando en guerras reales causando sufrimiento real, pero la razón subyacente es la motivación de ganar poder manteniendo una ventaja tecnológica. Este tipo de razonamiento no es nuevo en la historia.

Si hablamos de IA y sociedad y observamos el orden internacional, también debemos recordar que la etapa del «Viejo Oeste» de la tecnología no es eterna. Toda tecnología puede tener efectos secundarios adversos que planteen nuevos retos a la hora de enmarcarlos y de hacerles frente. Conducir un coche, ahora requiere un permiso de conducir, y estamos acostumbrados.

En cuanto a la IA, la UE no es la única impulsora de normativas. China empieza a aplicar más normas, y lo mismo ocurre en Australia, Malasia y otros países, con respecto al uso de los medios sociales por parte de los menores.

Sin embargo, las normas y los reglamentos también se aplican al puro poder del mercado. Los EE.UU. han sido excelentes en el desmantelamiento de monopolios, como en el caso de las sociedades ferroviarias, las petroleras y las telefónicas. Hasta ahora, no han conseguido acabar con el poder de los gigantes tecnológicos, aunque haya un debate en curso sobre ese tema. Desde el punto de vista del director general de un gigante tecnológico, un país no necesitaría

un presidente conforme a las elecciones democráticas, sino un director general que se encargue de las decisiones necesarias. De nuevo, no debería sorprendernos que personas como Elon Musk, junto con Peter Thiel y otros, aboguen por una fuerte influencia política de plataformas como X (antes Twitter). Lo mismo ocurre en el contexto de gobiernos autoritarios como en Rusia, China o Corea del Norte. La libertad de expresión se ve entonces como una amenaza para el Estado, no como parte de la cultura democrática y la diversidad.

Los límites de la libertad de expresión en la esfera pública se aprecian en el caso de las noticias falsas y el discurso del odio. Como toda regulación propuesta será el espejo de una perspectiva determinada, es necesario debatir hasta encontrar un compromiso que pueda ser aceptable para todas las partes interesadas.

Esto nos lleva del ámbito de actuación nacional al internacional y global. De hecho, las noticias falsas y la incitación al odio también pueden utilizarse como medio de propaganda y de guerra cibernética. Sin embargo, no se pueden imponer nuevas normas y reglamentos. Deben desarrollarse en un proceso de intercambio de argumentos de muchas partes.

Normalmente, la investigación de los hechos es un primer paso útil en la construcción de un nuevo orden. Un segundo paso puede ser la creación de instituciones nuevas y dignas de confianza. Por último, los acuerdos internacionales pueden allanar el camino hacia un futuro más pacífico.

A nivel nacional y/o regional, podríamos debatir algo así como una Agencia de Datos e Inteligencia Artificial de la UE que tuviera la tarea de recopilar y clasificar los tipos de uso tecnológico ambivalente de la IA. Desde el punto de vista de la sociedad civil global, puede llegar el momento en que necesitemos un Tribunal Digital de la ONU. La razón es sencilla: ¿qué debo hacer contra mi propio Estado en caso de violación de mis derechos civiles debido a un exceso de control autoritario? En ese caso, será necesaria una institución internacional.

Este punto es fácil de entender pero difícil de realizar, debido a la resistencia de los Estados que tendrían que aceptar un mínimo de control externo. No obstante, aunque las nuevas instituciones internacionales duren 30 o 50 años, o incluso un siglo, antes de hacerse realidad, es necesario abrir el debate para allanar el camino.

Veamos otros dos ejemplos: la guerra impulsada por la IA y la IA en el espacio. Las Convenciones de Ginebra de 1949 y los Protocolos Adicionales de 1877 y 2005 muestran en su esencia la capacidad de la humanidad para encontrar normas y reglamentos incluso en casos de conflictos armados. Hasta hoy, también el inconveniente es obvio: no existe una aplicación vinculante de tales acuerdos internacionales. La consecuencia es una tragedia, que da lugar a muchas violaciones y más bien al imperio del poder que al imperio de la ley. Pero la sociedad civil mundial funciona de alguna manera, ya que siempre habrá voces que utilicen los Convenios como referencia y que planteen la preocupación de tales violaciones.

De forma similar, necesitamos un nuevo Protocolo para el asesinato selectivo mediante drones asesinos manejados por IA (Taddeo, Floridi 2018). Existe una contradicción flagrante entre un Estado de Derecho con procedimientos bien definidos contra los criminales y la ejecución extrajudicial de personas por fuerzas militares incluso en territorio extranjero. Hemos visto casos de este tipo en la guerra entre Israel y Palestina a raíz de los atentados del 7 de octubre de 2023, extendidos tanto al Líbano como a Irán, donde se ha matado a miembros de Hezbolá utilizando aviones teledirigidos.

Además, existe una necesidad evidente de regulación global en el espacio. Actualmente utilizamos los satélites para muchos fines, desde la previsión meteorológica hasta la agricultura, desde las misiones espaciales hasta la guerra. No hay normas claras ni sanciones aplicadas en caso de abuso de poder por parte de una empresa privada o de un Estado.

Debatir sobre la IA desde la perspectiva de la sociedad civil mundial, en el mejor de los casos, puede tener un impacto en el establecimiento de la agenda. Las nuevas tecnologías no son neutrales. Se utilizarán en beneficio de los individuos, de las empresas, de los Estados, a veces más allá de los límites humanitarios razonables. La tarea de nuestra generación no es sólo el uso eficiente de la IA aprovechando sus increíbles ventajas para la gestión del conocimiento, la eficiencia de los procesos y la comunicación. Nuestra tarea consiste también en un proceso de búsqueda colectiva para superar la etapa «Viejo Oeste» del uso de la IA.

Desgraciadamente, no parece que haya un final para un mundo con múltiples conflictos. Sin embargo, es una noble tarea académica y política buscar formas y medios para un mundo en equilibrio, utilizando los mejores lados de la IA en beneficio y no en detrimento de los seres humanos en la sociedad civil global.

Los datos son poder (K. Crawford 2021, K. Schüller 2025). Depende de nosotros, los seres humanos, encontrar el equilibrio adecuado entre confianza y control. Por último, la ambivalencia de la IA no es simplemente una ambivalencia tecnológica. Es más bien la ambivalencia ética de la toma de decisiones humana (U. Hemel 2020, P. Kirchschräger 2021). Será tarea de la humanidad encontrar la forma de prohibir los abusos y sacar el máximo partido de la IA como un cambio de juego en la historia de la humanidad.

Prof. Dr. Ulrich Hemel

Traducido al español por Marius Rouger

Sources

Acemoglu, Daron.; Robinson, James A. 2019, *The Narrow Corridor: States, Societies, and the Fate of Liberty*. London.

Crawford, Kate. 2024 *Atlas der KI, Die materielle Wahrheit hinter den neuen Datenimperien*, München: C.H. Beck (=Atlas of AI, 2021: Yale University Press).

Hemel, Ulrich. 2020. Kritik der digitalen Vernunft. Warum Humanität der Maßstab sein muss. Freiburg/Br.: Herder (=Crítica de la razón digital, Lima: USIL 2023).

Hemel, Ulrich, 2022. Menschenrechte, Globalisierung, Wirtschaft: Eine sozioethische Betrachtung, in: Gunter Geiger, Ulrich Hemel, Sabine Schössler (eds.), *Globalisierung, Menschenrechte und Wirtschaft*, Paderborn 2022, 93-100.

Hemel, Ulrich. 2024. Digitale Risikokompetenz entwickeln, Ethik-Datenschutz-Inklusion und die Herausforderung ganzheitlicher digitaler Bildung, in: *Weiterbildung 2024*, Nr.4 (August 2024), 25-29.

Hemel, Ulrich. 2025. Der Index inklusiver Entwicklung als Handlungspfad zu menschenwürdiger Globalisierung und zu einen globalen Ökosozialfonds, in: *Jahrbuch Nachhaltige Ökonomie 2023/2024*, Marburg 2025: Metropolis-Verlag, 157-170.

Kirchschläger, Peter G. 2021. *Digital Transformation and Ethics, Ethical Considerations on the Robotization and Automation of Society and the Economy and the Use of Artificial Intelligence*. Baden-Baden: Nomos.

Schüller, Katharina. 2025. *Daten sind Macht, Kompetentes Entscheiden im Zeitalter von KI*, Frankfurt/M. und New York: Campus 2025.

Taddeo, Mariarosaria, Floridi, Luciano. 2018. Regulate artificial intelligence to avert cyber arms race. In *Nature*, Bd. 556, Nr. 7701, 2018, 296-298.