TopHic | dig deeper

June 2023

Regulating AI, a categorical imperative

Expert's opinion

AI Act: what the future holds

Carlo Giuseppe Saronni

Partner & Head of Legal Bernoni Grant Thornton

"Artificial intelligence could lead to human extinction". Thus warned just a few days ago Sam Altman, top AI expert and CEO of Open AI, the company behind ChatGpt, the platform that revealed the potential of artificial intelligence to the general public. Elon Musk also prophesised dramatic scenarios related to AI. 350 scientists working in the field of artificial intelligence recently issued this alarming warning: "mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war". The predictive Terminator movie saga will come to mind to the less young. In the 1984 James Cameron film, set in a near future, a global AI-based defence network known as Skynet gained self-awareness...

read moreo

Overview

Is the evolution of AI what we anticipated?

Gigliola Pirotta

Responsible for Labour Law Bernoni Grant Thornton

During an interview with Life in 1970 Marvin Minsky, one of the godfathers of artificial intelligence, predicted that in a few years' time we would have a machine "with the general intelligence of an average human being" (so-called general intelligence).Undoubtedly, if to date the idea of such an intelligent machine is a utopia based on the superiority of digital logic over human logic and on the deconstruction of space and time, the role of artificial intelligence...

read more

Focus on

Is Artificial Intelligence discriminatory? A question of ethics

Ilaria Giuseppina Penco

AI & Data Ethics Counsel Intesa San Paolo

Discrimination occurs when a person is denied an opportunity or a right due to a certain opinion based on groundless and inappropriate reasons. Sometimes, artificial intelligence systems are not impartial and show prejudices that result in discriminatory algorithm outputs against an individual or a group of individuals due to their race, gender, age, etc. From a scientific perspective, this phenomenon is known as "bias", i.e., a distortion caused by prejudice. This does certainly not concern only algorithms, but also human minds. Humans have cognitive bias, i.e., systematic distortions of judgment that, substantially...

read more

Supplemento Notizie flash n° 107 del 25 febbraio 1984

Grant Thornton

Overview

Is the evolution of AI what we anticipated?

Gigliola Pirotta
Responsible for Labour Law Bernoni Grant Thornton

During an interview with Life in 1970 Marvin Minsky, one of the godfathers of artificial intelligence, predicted that in a few years' time we would have a machine "with the general intelligence of an average human being" (so-called general intelligence).

Undoubtedly, if to date the idea of such an intelligent machine is a utopia based on the superiority of digital logic over human logic and on the deconstruction of space and time, the role of artificial intelligence in our society is a central issue to the current debate on the topic.

Artificial intelligence is actually a tool which can support human beings in their tasks and activities, as well as a system to predict future decisions, thus with a potential significant impact on our choices.

The debate on AI originates first of all from an understanding of what is meant by artificial intelligence and the mechanisms underlying its functioning.

As known, there is no definition of artificial intelligence which includes the various functions and applications of this scientific discipline, except the one specifying that its purpose is to "define or develop programs or machines with a behaviour that would be defined as intelligent if it were exhibited by a human being" (F. Rossi).

In the last few years, the reach and impact of artificial intelligence have expanded at a breathtaking pace due to the exponential growth of computer speed, to the huge increase in the amount of data available and to cloud computing, to the point that it has become the key driver of the "fifth industrial revolution", characterised by the interaction between intelligent machines and human beings.

These factors, together with the various capabilities of machines (planning, vision, expert system, natural language processing, robotics and speech), among which machine learning, have made AI forecasting increasingly accurate, thus leading to the application of AI within specific domains (so-called weak artificial intelligence): from medicine to finance, from manufacturing to education, from weapons to information, up to resource management.

One of the most evident areas in which artificial intelligence has made significant progress is natural language recognition and image recognition.

Deep learning algorithms based on artificial neural networks, with layers of simple computational nodes comparable to neurons, have shown an increasing accuracy in recognising objects in images and in processing human language in a natural way, showing human-level performance on various professional and academic benchmarks.

An example of this is ChatGPT, the new artificial intelligence chatbot made available freely to the public by OpenAI in November 2022, capable of answering questions, chatting with people and generating texts based on natural language text inputs, and powered by a Large Language Mode (LLM), i.e. a deep learning model trained on a large body of texts. In two months, the tool was used by 100 million users (USB), who could test the output of an artificial intelligence capable of generating knowledge from what learned from human beings, without the intermediation of an expert.

Other major progresses concern unsupervised learning, where AI algorithms acquire "raw" unlabelled data and, based on a simplified representation of their training dataset, learn to generate new texts and new creative contents, such as music and art.

GPT 4.0 (Generative Pre-trained Transformer), the model licensed by OpenAI in March 2023, showed us that thanks to neural networks, machines can write emails, blog posts or articles (copy.AI) but, in their applications, they can also generate multimodal content such as images (Midjourney), compose music (Soundraw.io), protect rights (DoNotPay), as well as create avatars (Anime AI).

Another sector in which artificial intelligence is refining its capabilities is that of predictive algorithms, which go beyond formal logic and the perimeter of their source code to generate a model using information extracted from data mining.

The validity of machine answers would not derive from prescriptive axioms, but rather from the ability to answer questions basing on models which could update themselves also through interactions with the surrounding environment and human feedback (reinforcement learning).

These machine learning algorithms can have applications in healthcare, finance and marketing and are transforming the way in which human beings express judgments. They require users to have an adequate mindset to spot possible algorithms errors deriving either from an imperfect application of the function built on test data (generalisation errors), or from the absence of some scenarios in learning data.

Despite its progresses and advantages, AI is not an unerring technology: its answers often need to be understood and interpreted.

Moreover, there are significant social risks related to the use of this technology which require consideration, without, however, envisaging catastrophic scenarios, such as those of a humanity reduced to automatons as depicted in the film Wall.E.

Among the main risks and damages deriving from the use of artificial intelligence are: an increased misinformation amplified by recommendation systems and by LLM, which can produce a convincing misinformation known as "hallucinations"; the strengthening of social inequalities due to the use of distorted learning data and results; the undermining of users' privacy, as AI models source data from the web or purchase them, but never provide information on which are used to train them.

The Center for AI Safety (CAS), a not-for-profit organisation based in San Francisco, identified eight large-scale risks related to the competitive development of AI, among which weaponization, power-seeking behaviour and value lock-in.

One of the main risks perceived is the automation of human work. AI may substitute a broad range of activities and professions, thus endangering the employment of millions of people in various industries. The economic gap could increase lacking adequate policies and measures to mitigate the negative impacts of technological unemployment.

Another perceived risk concerns the ethics of AI. AI systems are not as good as the data they are trained on: besides perpetuating existing biases or discriminating against certain groups of people, they could generate, even by deception, a "rogue" artificial intelligence (Yoshua Bengio).

Data privacy and security are additional major issues related to AI. Automated learning algorithms require huge amounts of data for their training, but the indiscriminate and unregulated use of such data, including synthetic data, and the difficulty to check their accuracy and relevance might lead to a violation of people's privacy in terms of control of the information provided and correctness of the decisions.

Moreover, the corruption or manipulation of input data sets and data security breaches may lead to an abuse of technology, as these are not immediately perceivable given the characteristics of the model, which does not work according to logical deductions and is not necessarily transparent.

In order to guarantee that the impact of artificial intelligence be fair, secure and sustainable for the society as a whole, the experts of the Organizations of the United Nations have recently identified three tools: regulation, increased transparency, and human supervision.

It is therefore a question of governing and addressing artificial intelligence basing on an anthropocentric approach to understand how human beings and machines could work together to the best of their abilities to face society's pressing challenges to expand and broaden human experience, rather than replicate or even substitute it.

It will be necessary to boost human creativity, curiosity and empathy (Paul McDonagh-Smith) and only when machine will come to epitomise the best of who we are and potentially could be, we will hold the keys to our future.

Experts' opinion

AI Act: what the future holds

Carlo Giuseppe Saronni
Partner & Head of Legal Bernoni Grant Thornton

"Artificial intelligence could lead to human extinction". Thus warned just a few days ago Sam Altman, top AI expert and CEO of Open AI, the company behind ChatGpt, the platform that revealed the potential of artificial intelligence to the general public.

Elon Musk also prophesised dramatic scenarios related to AI.

350 scientists working in the field of artificial intelligence recently issued this alarming warning: "mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war".

The predictive Terminator movie saga will come to mind to the less young.

In the 1984 James Cameron film, set in a near future, a global AI-based defence network known as Skynet gained self-awareness rebelling against humankind and causing a nuclear holocaust.

In the end, a humanity on the brink of extinction will be saved just thanks to some reconverted machines.

So, which future awaits us? Will machines win?

Meanwhile, the future is already here.

Artificial intelligence is progressively becoming part of our lives.

Without noticing, on a daily basis we are parts of processes which involve systems with various degrees of autonomy.

We receive films recommendations, we rely on analyses made by algorithms, we delegate increasingly important activities and with a social impact.

In this way, we reduce interactions with human beings and we expose ourselves to technologies whose functioning is mysterious.

Faced with a rapidly evolving and expanding

scenario, Countries have assumed different roles in the international arena.

While China and US are the main investors and driving forces behind technological innovation, the European Union aspires to take on the role or "ethical legal champion".

The outcomes of this intention are taking the form of a long-awaited provision, known as "Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts".

The provision was presented to the European Commission on 21 April 2021 and submitted to the vote of the European Parliament between 12 and 15 June 2023.

The AI Act is the world's first regulation on artificial intelligence.

The provision has a horizontal scope, as it aims at regulating artificial intelligence in all areas of potential application.

It deals with the allocation of responsibilities and with the protection of fundamental human rights, such as health, safety and other fundamental rights of citizens who interact with AI systems.

The long gestation of the AI Act does not depend only on red-tape delays, but also on the need of lawmakers to deal with very general issues involving science, philosophy, ethics and law.

Even the definition of artificial intelligence is quite a debated crucial issue, since identifying the items which fall within the scope of the provision means understanding what is liable to controls, limitations, penalties, and what is not.

Moreover, understanding which models need to comply with the law requirements provided under the new regulation is key to directing technological innovation.

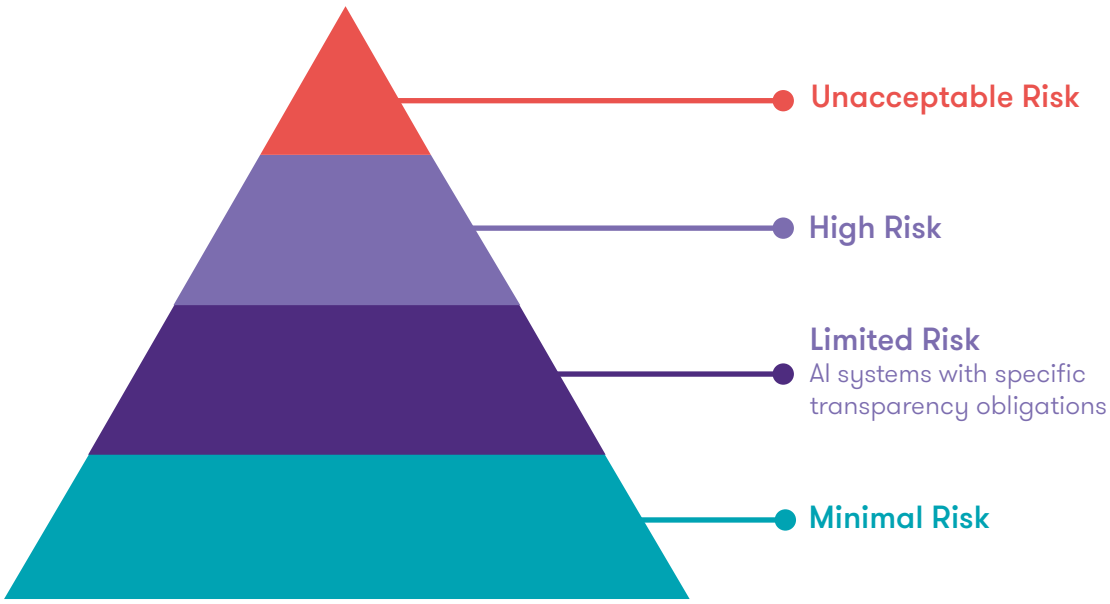
Indeed, the system which will elude the scope of application of the AI Act will attract greater attention from economic actors and will sometimes be preferred, to the detriment of those technologies which will have to comply with the strict requirements of the abovementioned regulation.

The AI Act classifies AI systems into four different levels according to their dangerousness and on the importance of the fundamental human rights to be protected: from minimal risk to limited risk, up to high risk and even unacceptable risk, where the relevant applications are banned.

The attribution of the highest risk level depends on the system's intended purpose and on its scope of application.

Systems dealing with lending or support to the administration of justice will be classified as high-risk based on the potential impact on people's lives and will have to comply with a series of requirements, among which an assessment of their impact on fundamental human rights.

Systems aimed at classifying people and used



for social scoring based on sensitive data (gender, race, ethnicity, religion, etc.), as well as predictive police systems based on profiling or sensitive data, or again emotion detection systems and face recognition systems for control purposes will be banned.

Considered as a whole, this regulation is an example of interaction between technology and law.

The trade-off between protecting rights and the ethical scenario and driving technological innovation is a topic of major interest for the main global players.

Open AI, for example, declared that should the regulation proposed by the European Parliament be too burdensome, it is ready to suspend its services in Europe.

The European institutions, for their part, replied they are not willing to yield to the blackmail of the big industry players.

The provision thus needs to take into account very fragile political-economical balances while trying to propose approaches in line with the European set of values, without putting off major players.

The AI industry actually has huge entry barriers given the infrastructure costs and the availability of computing power, so that only a few companies play a key role for the development of technologies.

On 31 May 2023, the Vice-President of the EU Commission Margrethe Vestager and the US Secretary of State Antony Blinken announced that a joint voluntary code of conduct on AI between the EU and the US will be introduced in the coming weeks, that will be open to the companies in the industry and will anticipate the norms included in the AI Act.

Moreover, the European Union is promoting initiatives such as LAION Open-Assistant, a research project financed by the German government aimed to create an "open-source Chat", showing that the promotion of technological innovation is possible in an ethical and legal context that cares about the impact of artificial intelligence systems on individuals' lives.

Europe is taking action, other countries will follow. For the moment, we can still have good possibilities to succeed, but this issue must be carefully addressed, as the other issues that have been neglected for a too long time, such as environmental pollution, global warming, overpopulation, etc.

Focus on

Is Artificial Intelligence discriminatory? A question of ethics

Ilaria Giuseppina Penco
AI & Data Ethics Counsel Intesa San Paolo

Discrimination occurs when a person is denied an opportunity or a right due to a certain opinion based on groundless and inappropriate reasons. Sometimes, artificial intelligence systems are not impartial and show prejudices that result in discriminatory algorithm outputs against an individual or a group of individuals due to their race, gender, age, etc.

From a scientific perspective, this phenomenon is known as "bias", i.e., a distortion caused by prejudice. This does certainly not concern only algorithms, but also human minds.

Humans have cognitive bias, i.e., systematic distortions of judgment that, substantially, derive from two different sources: the first one is biological, while the second one is the result of the cultural and social context in which an individual grows up and lives.

We can affirm that the first source is "infrastructural" and the second one is "informational".

For example, the order in which options are presented conditions human decisions: the human brain "prefers" some things over others because they are presented in a certain position compared to others. An example is the way products are presented on supermarket shelves or the way movies are presented on platforms. This is "infrastructural" bias.

Typical "informational" bias is gender bias. If we are used to see men at top positions, we could be inclined to incorrectly believe that women cannot reach leadership positions.

Artificial intelligence systems are potentially subject to the same distortions or bias that condition humans.

An artificial intelligence system being used to suggest the most suitable applicant for a vacancy could provide unappropriated suggestions due to infrastructural reasons related to the algorithm, which could have some defects causing a certain characteristic to be under- or overestimated. Otherwise, it could provide a wrong or discriminatory suggestion due to informational reasons, i.e., reasons related to the data used to "train" the system.

In fact, data could present distortions just like the environment in which humans grow up: in fact, they are a product of such environment. Hence, the expression "garbage in garbage out": if data are of poor quality, algorithm predictions will be of poor quality, too, and will generate errors or distortions.

A typical area in which bias can be found is that of equality between men and women or between people of different races or religions, because our economic and cultural system has been based for millenniums on discriminatory behaviours, which are often supported by legal systems.

If female applicants in a certain company have always received a lower remuneration than male ones, an algorithm used by an artificial intelligence system that was trained on those data will be inclined to propose women a lower average remuneration. This kind of distortion is called historical bias and, in many cases, it is the result of a social inequality in training data.

The emerging Artificial Intelligence Act (please refer to Carlo Giuseppe Saronni's article in this TopHic issue) should also deal with this kind of bias.

In fact, art. 29 (a) of the text proposed by the Parliament provides for an impact assessment on fundamental rights, based on which the system user should make an assessment on high-risk systems aimed to identify potential breaches of fundamental rights, including the right to non-discrimination.

As said, bias can also depend on a poor representativeness of data.

In general, the more the data received by an artificial intelligence system on a certain area or matter or on a certain population, the more solid and reliable its prediction and capacity to represent it properly.

Consider, for example, racial discrimination: if few data are available for a certain race, the system prediction on that race will not be fully reliable since the system is trained on few information.

Then, if characteristics (e.g., low income) are constant for that race, the system will associate a certain race with a low income also in opposite circumstances, thus damaging persons belonging to that group.

To solve the low representativeness of data, the regulation should provide under art. 10 (3), that "data must be representative" of the population which the algorithm has effect on.

The rationale behind the regulation is commendable.

However, it is difficult to concretely comply with this provision.

In fact, data available to the algorithm will always depend on the reference population up to that moment and, therefore, it will sometimes be difficult to avoid some cases of misrepresentation.

Lastly, some kinds of bias depend on the type of data used.

There are some human characteristics that can hardly be represented.

Intelligence, moral sense, empathy, the ability to conciliate people are aspects that can hardly be represented by measurable data and, subsequently, evaluated, for example, by a personnel recruitment algorithm.

In this case, again, the output provided by the machine could be discriminatory or non-efficient for the user, since it could lead to the selection of a less deserving applicant only because fundamental qualities are not captured by data and therefore are not considered.

One last example.

Intelligence is a complex of mental faculties that distinguishes humans from animals.

IQ can be measured. However, IQ regards a very limited aspect of human intelligence, i.e., logical abilities (elementary inferences that involve short-term memory) and spatial visualization (most of all, rotation and pattern recognition).

These abilities, either innate or trained, certainly favour those professions that are more based on spatial reasoning, such as mathematicians and physicists, who have a higher preparation in geometry. These individuals will be favoured by an artificial intelligence algorithm that measures human intelligence based on IQ.

In fact, even if the machine is very sophisticated, it needs an element to be measurable, in order to give it a value.

In this sense, IQ is a measurable, established, and privileged parameter to quantify intelligence.

Nonetheless, this parameter cannot properly capture the whole dimension that it aims to measure, i.e., intelligence, since measurable data could weight more than other important – but non easily measurable – personal qualities.

Therefore, the evaluation of intelligence based on IQ represents a synthetic but not exhaustive solution to represent the characteristics of an applicant, because the system could exclude deserving individuals due to the impossibility for its impossibility to process hardly measurable aspects of intelligence.

Clearly, these are very sensitive issues.

Data scientists cannot be left alone in considering these problems but should rather be supported by jurists and philosophers who can help them identify the possible bias in the artificial intelligence system and the best technologies to mitigate them.

Luckily, large companies are recruiting these professional figures, who should help data scientists organize their work.


We don’t predict the future. We help you shape it.
BGT-GRANTTHORNTON.IT
Grant Thornton