Senato della Repubblica

Ufficio Valutazione Impatto
Impact Assessment Office

FOCUS

The challenges of artificial intelligence

Governance, accountability and privacy in the age of systems that display intelligent behaviour

July 2024

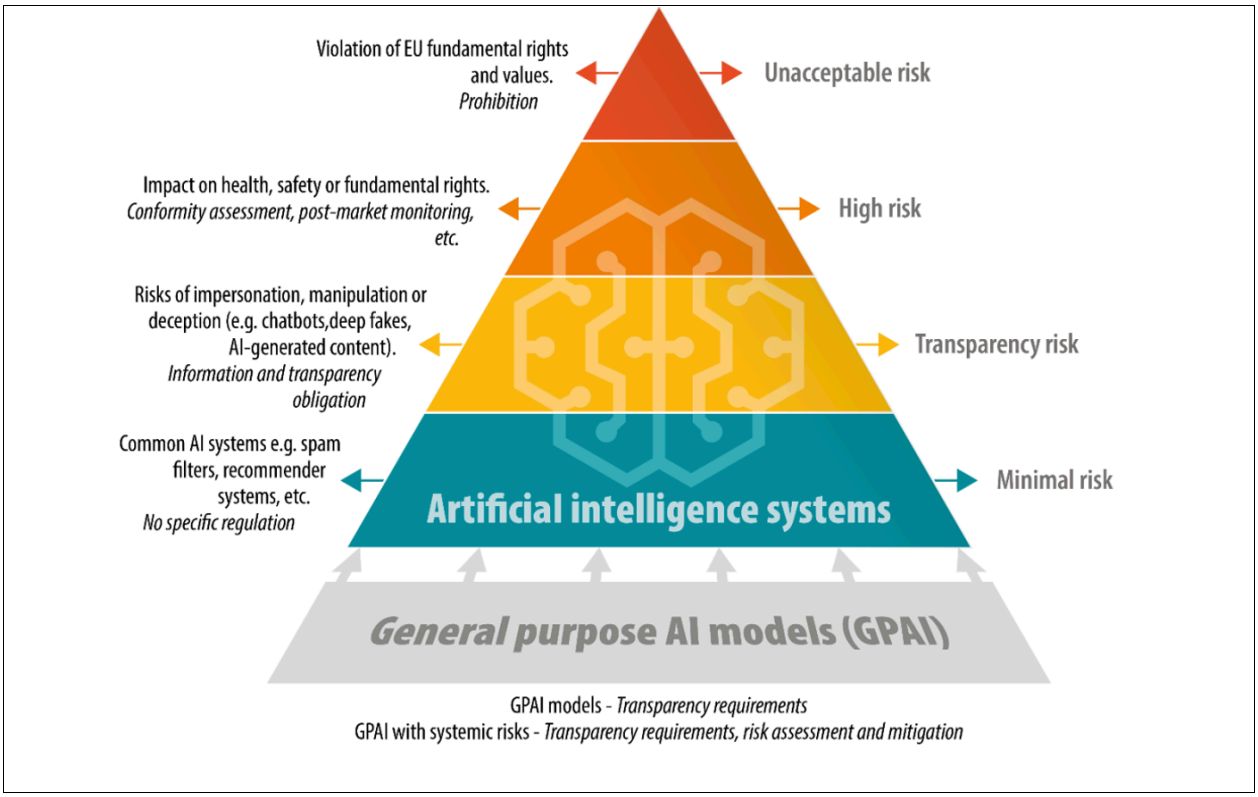
*According to the definition adopted by the European Commission in 2018, artificial intelligence (AI) is defined as "**systems that display intelligent behaviour** by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals". These systems, which are characterised by rapid and pervasive deployment, **impact all human activities** with technological, legal, economic and social implications.*

*Many countries have therefore begun to **regulate** aspects and consequences of AI use, but faced with the speed at which technologies evolve, formulating comprehensive regulations appears to be a **challenge**. However, **many issues remain open,** such as the protection of privacy and personal data, the **protection of intellectual property** and generative AI products, the risk of **profiling judges** and the advent of '**predictive justice**'.*

The starting point

During its plenary session on 13 March 2024, the European Parliament finally adopted the Proposal for a Regulation laying down harmonised rules on artificial intelligence (*Artificial Intelligence Act*), which will enter into force 24 months after its publication in the Official Journal of the Union.

Figure 1. Risk levels in the use of AI according to the European Commission



Source: European Commission

The scenario

The term 'Artificial Intelligence' (AI) was coined at the *Dartmouth Summer Research Project on Artificial Intelligence* conference held in **1956** at Dartmouth college in New Hampshire, which laid the foundation for the development of AI as we know it today.

The work of Alan Turing played a key role there. In 1950, he published an article in *Mind* entitled *Computing machinery and intelligence*, illustrating what later became known as the '**Turing test**', a **criterion for determining whether or not a machine exhibits intelligent behaviour**: a machine can be considered intelligent if its behaviour is indistinguishable from that of a person.

In **1958**, psychologist and engineer Frank Rosenblatt worked on the design of a machine 'capable of perceiving, recognising and identifying its surroundings without any training or control by humans': the *Mark I Perceptron*, the **first model of *artificial neural network***, modelled on the human neuronal system.

In the same year, John McCarthy followed a different approach: instead of having artificial neurons work and produce a numerical output, **human knowledge was encoded in logical rules through the programming language LISP** (*List Processor*).

This approach was then widely developed in the **1980s** in Japan and based largely on the work of Edward Feigenbaum. The American computer scientist had introduced the field of **'expert systems' that mimicked the decision-making process typical of human beings**.

A further step forward came in **2012** when the *AlexNet* 'convolutional neural network' model proposed by Alex Krizhevsky and Ilya Sutskever marked a significant advance in **automatic image recognition**.

In **2020**, by developing an AI programme called *AlphaFold*, Google's DeepMind division managed to solve one of biology's greatest challenges: determining the 3D structure of a protein from its amino acid sequence. The **ability to predict protein structures** has been a revolution, offering a powerful tool to support drug and therapy research.

On 30 November **2022**, **OpenAI** - a US artificial intelligence research lab - launched the *Chat*

Generative Pre-Trained Transformer (**ChatGPT**). It represents a *Large Language Model* (LLM) that uses ***transformer*** **neural networks to generate texts in a coherent and contextually relevant manner**. The GPT architecture is characterised by a training phase on huge amounts of textual data from the Internet. The model is periodically updated on an increasing amount of data.

The giant **Meta** has also invested time and resources in AI over the years. In 2013, it founded the Facebook AI Research (FAIR) group, which led to the introduction of *BlenderBot3* and *Galactica*, which, however, did not produce satisfactory results. It was only in July **2023** that Meta was able to compete with other big tech, with the introduction of the LLM called Llama2*,* developed in collaboration with Microsoft and available under an *open source* license.

Google has launched several versions of its own LLM on the market, including *Bard* in May 2023 and *Gemini* in February **2024**. One of the distinctive features of Google's models is the **availability of constantly updated training data**. For instance, the ChatGPT -4o knowledge cutoff date is October 2023, and thus all questions concerning later events cannot be answered. With the introduction of *Gemini,* on the other hand, Google declared that it wanted to reinvent its products ecosystem directly integrating artificial intelligence, with functions also available offline.

Why do we need regulation?

Many countries have embarked on a path to regulate various aspects and consequences of the use of artificial intelligence. Faced with numerous sectors that could be impacted by new technologies and, at the same time, faced with the speed at which these evolve, the formulation of comprehensive regulations appears to be a challenge, but also **a necessary action for the protection of citizens and society** at large. Indeed, the G7 leaders, in the press release issued at the end of the summit that took place between 13 and 15 June 2024 in Italy, reaffirmed their intention to cooperate on these issues, with the aim of pursuing an inclusive and 'human-centric' digital transformation that supports economic growth and sustainable development, in line with shared democratic values and respect for human rights.

Europe's priorities

In 2018, the European Commission set up a High Level Expert Group on AI, which in April 2019 published the *Ethics Guidelines for Trustworthy AI*. The document, in addition to suggesting **guidelines** for AI based on legality, ethicality and robustness, describes **examples of 'concerns raised by AI**': e.g. the automatic recognition and identification of people through the use of biometric data; the need for users to always be able to know whether they are interacting with a machine, in order to avoid 'consequences such as attachment, influence or reduction of the value of being human'; the evaluation 'by score' (so-called social scoring) that jeopardises the autonomy and freedom of citizens, undermining the principle of non-discrimination.

In its subsequent *White Paper on Artificial Intelligence,* the Commission elaborated on the risks, considering that the use of AI in certain fields could '**undermine the values on which the Union is founded and cause violations of fundamental rights**, including the rights to freedom of expression and assembly, human dignity, non-discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation [...], the protection of personal data and private life or the right to an effective judicial remedy and a fair trial, as well as consumer protection'.

The European Commission identified the areas considered to be 'unacceptable risk' and 'high risk' (Figure 1) and presented **a proposal for a regulation, also known as the *AI Act,*** on 21 April **2021**. This is the most complex text

globally concerning AI. **The regulation was finally adopted** in the plenary session of the European Parliament on 13 March 2024 and will enter into force 24 months after its publication in the Official Journal of the EU.

Sectors at 'unacceptable risk'

Some particularly harmful uses of AI contravene EU values because they violate fundamental rights and have therefore been banned by the *AI Act*:

- **Social scoring** for public and private purposes;
- **Exploitation of people's vulnerabilities**, use of subliminal techniques;
- Real-time **remote biometric identification** in publicly accessible spaces **by law enforcement agencies**, subject to certain exceptions (see below);
- **Biometric categorisation** of natural persons based on biometric data to infer race, political opinions, trade union membership, religious or philosophical beliefs or sexual orientation. Filtering of data sets based on biometric data within the scope of law enforcement will still be possible;
- **Individual predictive policing**;
- **Recognition of emotions in the workplace and in educational institutions**, unless for medical or safety reasons (e.g. monitoring a pilot's fatigue levels);
- **Untargeted collection of facial images from the Internet or CCTV cameras** to create or extend databases.

High-risk sectors

- Some **critical infrastructures, e**.g. in the areas of road traffic and the supply of water, gas, heating and electricity;
- **Vocational education and training**, e.g. to assess learning outcomes and guide the learning process and control monitoring;
- **Employment, worker management and access to self-employment**, e.g. to publish targeted job advertisements, analyse and filter job applications and assess candidates;
- **Access to essential public and private services and benefits** (e.g. health care), creditworthiness assessment of individuals, risk assessment and pricing in relation to life and health insurance;
- Some **systems used in the areas of law enforcement**, border control, administration ofjustice and democratic processes;
- Evaluation and **classification of emergency calls**;
- **Biometric identification**, categorisation and emotion recognition systems (outside prohibited categories).

All other AI systems, which are considered 'minimal risk', **can be developed within the limits of existing legislation**. They will, however, have to meet **minimum requirements in terms of transparency** to enable users to make informed decisions.

What happens elsewhere?

United States. On 30 October 2023, President Joe Biden signed an Executive Order on AI, addressed to government agencies, which aims to set new standards for security and privacy protection, for the promotion of civil rights, for the defence of consumers and workers, and for the promotion of innovation and competition. In November, Vice-President Kamala Harris announced further initiatives, such

as the creation of the *United States AI Safety Institute* (US AISI), with the task of formulating guidelines and identifying tools and standards to mitigate any risks.

United Kingdom. In February 2024, a number of 'cross-sectoral' principles were identified for regulators to interpret and apply, each in their own area, with respect to innovations resulting from the application of AI.

Switzerland. In November 2023, the Federal Council assigned the Federal Department of the Environment, Transport, Energy and Communications (DETEC) the task of exploring options for regulating AI: a report including possible EU-compatible regulatory solutions is expected to be published by the end of 2024.

China. Generally speaking, Chinese regulation does not cover all fields of AI use, but it is aimed at specific applications such as the use of algorithms (*Algorithm Recommendation Regulation*) and the use of generative AI.

Council of Europe. It is addressing the topic in numerous committees, including the *Committee on Artificial Intelligence* (CAI), and on 17 May approved the *Framework Convention on the development, design and application of Artificial intelligence*.

United Nations (**UN**). In October 2023, the UN established the Artificial *Intelligence Advisory Body*. The primary objective of this body is to consolidate a 'scientific consensus' on the risks and challenges of artificial intelligence and identify ways to make the most of it.

From judges profiling to generative art: how artificial intelligence is creating new scenarios (and new legislative challenges to deal with them)

1. Are privacy and personal data at risk?

A particularly critical aspect of the development and spread of artificial intelligence concerns the **protection of privacy and the use of personal data.** The *AI Act* contains precise provisions regarding uses of AI that are deemed to pose an 'unacceptable risk': for example, untargeted scraping of facial images, recognition of emotions in the workplace and in schools, so-called 'social scoring', and biometric categorisation to deduce sensitive data such as sexual orientation or religious beliefs are prohibited. The EU provisions for the protection of personal data remain fully applicable.

A debate has also arisen in **Italy** on the need to protect citizens' privacy when using AI. In March 2023, the *Garante per la Protezione dei Dati Personali* (GPDP) ordered OpenAI **to temporarily restrict the processing of Italian users' data**, effectively blocking access to ChatGPT. The US company then undertook to implement some changes following the measures indicated by the GPDP.

In December 2023, the GPDP launched an investigation into the online collection of personal data for the training of AI systems. Pending its conclusion, in January 2024, the *Garante* **notified breaches of data protection law** to OpenAI, giving the company 30 days to submit its counterclaims.

Finally, on 8 March, the GPDP opened **an investigation into *Sora***, OpenAI's artificial intelligence model able to create videos from short text instructions. The investigation aims to understand the implications that the tool might have on the processing of users' personal data.

2. AI in court. What future for 'predictive justice'?

By 'Predictive justice' is meant the **prediction of the outcome of the judgement** by both the judge and the litigants, including lawyers.

From the judges' point of view, AI systems could be effective in handling **cases of a minor and 'repetitive' nature**: it would be possible to reduce trial times by devolving the burden of an initial 'judgement' to electronic processors, whose decision could in any case be appealable and reviewable on the basis of the principle of 'man-machine complementarity'.

From the lawyers' point of view, on the other hand, the new technologies could prove useful not only with regard to the **assessment of the likelihood of success in a case**, but also in the area of **'precedent research'** - both jurisprudential (relevant, in particular, in Common Law countries) - and with regard to the **'personal predispositions' of the individual judge**.

This aspect is not ignored by European legislation, which includes among high-risk AI systems those 'intended to be used by or on behalf of a judicial authority to interpret facts or law and to apply the law to a concrete set of facts'.

Projects - public and private - on **predictive justice have already been launched in several European and non-European countries**. **In Italy, experiments are being carried out** by some courts, also in collaboration with university laboratories.

In December 2023, the **public merit database** that collects judgments, orders and decrees in civil law matters issued by courts and courts of appeal since 1 January 2016 became operational. Among the published data are the **names of the judges** who pronounced the measures.

This could increase the risk of data being used for predictive purposes, even to the point of influencing the decisions of judges themselves: by means of special algorithms, it might in fact be possible to identify 'tendencies' or 'propensities' of judges in the application of the rules and, by means of past data, 'predict' or even influence future decisions.

This is the so-called **'profiling' risk**, which has already been the subject of regulatory intervention in some jurisdictions. The topic has also been the subject of reflection in the **Council of Europe**, which has pointed out that, through these systems, there could be a **risk of violation of the principle of due process** in some jurisdictions - for instance, through the practice of so-called 'opportunistic choice of court'.

Italy chose to limit, on the data made available in the merit database, the possibility of performing certain classification, evaluation, comparison and profiling functions.

3. Intellectual property. Who owns the copyright?

The generative AI models developed in recent years have the capacity to produce content of different kinds - written texts, images, videos - in fact 'learning' from already existing content and creating new content. Critical issues have thus arisen with respect to the **material used for so-called model *training*.**

- **Protection of the authors' product**: the training process of generative AI is based on research from web sources (often public, but also private) such as books, articles, photographs, songs and videos. Data are usually copied and re-processed on the basis of the algorithms used by the AI, depending on its purpose. This massive collection of data is likely to include copyrighted material. Today's challenge therefore appears to be to identify general and certain criteria to determine the limits within which generative AI can go without infringing copyrights and patents.

- **Protection of the product created through generative AI tools**: is there a way to protect the creator who used the AI tool, or should AI itself have a form of legal protection? Already in 2015, the European Parliament stated it would not be appropriate to seek to impart legal personality to AI technologies, also considering the associated risk of a negative impact on 'human creators'. On the other hand, in a world where the use of AI in the artistic process is increasingly common, the lack of legal protection for products could leave a protection gap to the detriment of the artists who helped to develop them. At the moment,

the issue, as complex as it is controversial, remains unanswered.

What ChatGPT thinks

How can ChatGPT, or other generative language models, violate intellectual property rights? Here are some scenarios identified by ChatGPT-4o.

- **Reproduction of protected texts**: if ChatGPT reproduces copyrighted texts in whole or in part without permission, it may infringe copyright. This may occur if replies include significant excerpts of literary works, articles, books, or other protected material.
- **Generating derivative content**: creating content that is derivative of protected works may constitute infringement. For example, if ChatGPT generates a story, poem or song that is clearly based on a protected work, it may infringe the rights of the original author.
- **Use of registered trademarks**: Using registered trademarks without authorisation, in a way that confuses consumers or in a context that may diminish or damage the reputation of the trademark, may lead to infringement. For example, if ChatGPT uses a trademark in an inappropriate or misleading manner, it could be considered a violation of trademark rights.
- **Reproduction of protected images**: when ChatGPT depicts or generates images based on copyrighted works of art, photographs or other visual materials without the permission of the rights holder, it may infringe intellectual property rights.
- **Sharing of source code**: Reproducing software source code that is copyrighted or has been published under a license that does not allow unauthorised sharing, constitutes infringement. ChatGPT may, for instance, reproduce parts of protected code in response to technical requests.

To minimise the risk of intellectual property infringement, it is important that ChatGPT users and developers use the template responsibly, respecting copyright, trademark and patent laws.
In addition, developers can implement security measures and filters to prevent the template from generating content that might infringe intellectual property rights.

Towards multilevel governance

The guidelines at European level aim to maintain a market surveillance system at national level and a more centralised supervision approach for general-purpose artificial intelligence (GPAI) models. To this end, the **Office for AI** was set up to collaborate with the scientific community.

The proposed governance structure envisages a more significant role for the **European AI Committee**, which will see its responsibilities extended to ensure greater coordination among the Member States.

There will be two new advisory bodies: a **group of independent experts** that will provide technical advice to the AI Office and market regulators, with a crucial role in identifying potential risks for GPAI models, and an **advisory forum** that will be a feedback channel for the Commission and the Council, ensuring a balanced representation between industry, start-ups, SMEs, civil society and academia.

Each EU Member State is required to establish or designate **a notifying authority** and **a market surveillance authority** as competent authorities, which will have to ensure objectivity and effectiveness in the application of the European regulation.

The Italian strategy for AI

At the event '*Artificial Intelligence for Italy*', which took place on 12 March in Rome, the government provided advance information on the National Strategy for AI: **the Council**

presidency will play a central role and **AI will be considered as a matter of national security and strategic interests,** thus also resulting in a **revision of the golden power regulation.**

A crucial aspect of the strategy is the focus on the world of education, with the aim of identifying creative talent at an early stage, preventing the so-called 'brain drain' and stimulating the development of national competences in the field of AI. **A public agency is envisaged**, which, in addition to monitoring the implementation of the strategy, will have supervisory functions and will be able to impose sanctions, following the principles of the European regulation.

With regard to the funding, the strategy envisages the creation of a foundation and the use of **resources to promote research and development**, as well as **support for start-ups and high-tech companies** with innovative projects.

Senate activities

The advent of AI is also posing unprecedented challenges and opportunities in the parliamentary context. Since the XVIII legislature, the Italian Senate has promoted a wide range of initiatives - **expert hearings, discussion of bills, international conferences, fact-finding and specialised publications** - with the aim of fostering the responsible adoption of this technology.

The activity has been intensifying In the current legislature, **out of 1,095 bills** presented in the first 20 months, **15 concern various aspects of AI**. Among them, DDL No. 1146, *Provisions and Delegation to the Government on Artificial Intelligence*, addresses five specific areas: national strategy, national authorities, promotion actions, copyright protection and criminal sanctions.

This dossier

It illustrates the origin and evolution of AI systems; deals with the most significant approaches to their regulation; analyses the responsibilities arising from the use of AI with regard to copyright protection, predictive justice and privacy; and illustrates the governance models and insights carried out by the Senate.

The study was carried out by
GIANPAOLO ARACO
SABRINA AURICCHIO
NICOLÒ DE SALVO
FEDERICA IZZO
ROBERTA MAGLIO
Senate of the Republic

Focus by
IMPACT ASSESSMENT OFFICE
Senate of the Republic
uvi@senato.it