

EURISPES NOTE

Research in progress on the Digital Codes of Ethics

May 2022

ETHICS - DIGITAL

**DIGITAL CODES OF ETHICS WITH LEGAL FORCE
AN INTERNATIONAL COMPARISON
(CANADA - USA – RUSSIA - EU)**

Specific initiatives in Italy and Germany

Introduction

The need to have Digital Codes of Ethics as an essential basis for better guiding and managing the complex processes of the ongoing digital revolution is widely felt at an international and national level, particularly concerning the use of Artificial Intelligence (AI) and of data (Big Data). In this regard, we find numerous documents and recommendations approved by public institutions and, in many cases, defined and taken on a voluntary basis by private operators themselves. In general, the objective pursued is to maximize AI's benefits and minimise the potential and real damages to individuals and communities.

However, there are still very few cases in which the initiative to promote and enforce Digital Codes of Ethics has been translated into a binding legal and regulatory instrument to be applied in society, the economy, and especially in the world of work. As of 2022, only three countries have officially adopted such tools with the force of law: **Canada, the United States, and the Russian Federation.**

Concerning the **European Union**, it should be noted that, after a preparatory work started in 2017, the Union presented an organic proposal on 21 April 2021, the “Artificial Intelligence Act” (AIA), with the aim to translate it into European law. It is foreseeable that the path to approval of this legislative act will take a few years.

This research provides a comparative analysis of the regulatory frameworks in digital ethics approved and currently in force in Canada, the United States, the Russian Federation, and of the EU document whose legislative process has begun. The Annex illustrates the specific initiatives promoted on the subject in **Italy** and **Germany**. This research aims to offer a cognitive contribution as a basis for accelerating the adoption of similar organic instruments of ethical regulation by other states worldwide, in Europe and Italy.

CANADA

Legislation on ethics in AI and automated decision-making system

Canada's legislation on digital ethics

Reference is made to two main acts:

- a) the *Advisory Statement on Human Ethics in Artificial Intelligence and Big Data Research* – 2017
- b) the *Directive on Automated Decision-Making (DADM)* - 2019

To understand the value of these two legislative acts, we should be borne in mind that Canada has always had a leadership position in AI, which is widely recognised internationally.

The Pan-Canadian Artificial Intelligence Strategy - 2017

In 2017, the country launched its \$125 million '*Pan-Canadian AI Strategy*,¹ becoming the first government to promote a national AI strategy. The strategy was based on a partnership between the *Canadian Institute for Advanced Research (CIFAR)* and three Canadian centres of excellence: the *Alberta Machine Intelligence Institute (AMII)* in Edmonton, the *Vector Institute* in Toronto, and the *Montreal Institute for Learning Algorithms (Mila)* in Montreal. The objectives it aimed to achieve were:

- Attract and support world-class AI researchers.
- Promote a collaborative AI ecosystem by establishing interconnected networks of scientific excellence among Canada's three major AI centres: Edmonton, Montreal, and Toronto.
- Promote national AI initiatives by supporting a national research community with training programs, workshops, and other collaborative opportunities.
- Understand the social implications of AI by developing global thought leadership on the economic, ethical, political, and legal implications of advances in AI.

The “Advisory Statement on Human Ethics in Artificial Intelligence and Big Data Research” - 2017

The same year, the *National Research Council of Canada (NRC)* published the "*Advisory statement on human ethics in artificial intelligence and big data research*."² This statement is binding for federal agencies and highlights ethics and respect for human dignity as an essential part of the excellence in sector research. With this declaration, the NRC pledged to comply with a series of requirements in scientific and engineering research activities and also stated its willingness to engage research staff and promote related projects only with the approval of the *NRC Research Ethics Board (NRC-REB)*, the body responsible for the ethical evaluation of research. The principles set out in the declaration for such an evaluation are as follows:

- **Protect privacy and personal information:** i) take extraordinary measures in research design and planning to protect the privacy of individuals and personal information about them, ii) recognise the right of individuals to access their personal information collected and stored by the NRC.
- **Ensure discreet and genuine consent:** (i) involving in research activities only those individuals who express free and informed consent to the use of their personally identifiable

¹CIFAR, Pan-Canadian AI Strategy - <https://cifar.ca/ai/>

²NSF Canada, Advisory statement on human ethics in artificial intelligence and big data research (2017) - <https://nrc.canada.ca/en/corporate/values-ethics/research-involving-human-participants/advisory-statement-human-ethics-artificial-intelligence-big-data-research-2017>

information; (ii) obtaining the consent of the participant in research projects for the use of personal information in a clear, unambiguous manner that is not conditioned on the acceptance or otherwise of any form of limitation, deprivation, control, as is the case, for example, in the purchase of certain products or services; (iii) research that relies solely on the secondary use of non-identifiable information does not require the consent of the participant; but in any case the approval of the NRC-REB is required.

- **Preserve human and legal rights:** i) ensure that consent to the use of personal data or in any case to participate in the research is not conditional and does not include any declaration that, by giving consent, participants renounce fundamental human rights or any right of legal remedy in the event of research-related harm, ii) respect the rights of affected communities as a whole when, for example, the data and information studied have implications for a specific community such as an indigenous people, iii) take special measures to protect the rights of children and other vulnerable individuals.

The “Directive on Automated Decision-Making (DADM)” - 2019

In 2019, the Government of Canada unveiled the *Directive on Automated Decision-Making (DADM)*³ and the *Algorithmic Impact Assessment (AIA)*, the accompanying tool of the directive for the assessment of algorithmic impact. The DADM, which went into effect on April 1, 2020, and the AIA are designed to guide the adoption and use of automated decision-making at the federal level. The two key definitions of the Directive of Artificial Intelligence and Automated Decision Making (ADM) are contained in Appendix A:

- **Artificial Intelligence:** information technology that performs tasks that would generally require biological brain power to be performed, such as making sense of spoken language, learning behaviors, or solving problems.
- **Automated decision-making system:** includes any technology that assists or replaces the judgement of human decision-makers. These systems draw from statistics, linguistics, and computer science and use techniques such as rule-based systems, regression, predictive analysis, machine learning, deep learning, and neural networks.

The DADM applies to statistical systems, tools or models used to recommend or make life-or-death choices. The requirements of the directive are linked to fundamental principles of administrative law such as transparency, accountability, legality, and procedural fairness and are divided into five categories:

1. Carrying out an impact assessment
2. Transparency
3. Quality assurance
4. Recourse
5. Reporting

The Directive requires an algorithmic impact assessment for any automated decision-making system, including the impact on the rights of individuals or communities. The purpose of the Directive is set out in Section 4, which states that:

- The directive's objective is to ensure that automated decision-making systems are employed to reduce risk to Canadians and federal institutions and lead to more efficient, accurate, consistent, and interpretable decisions under Canadian law.

³Government of Canada, Directive on Automated Decision-Making, 2019 - <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>

- The expected outcomes of the directive are that (i) decisions made by federal government departments are data-driven and meet the requirements of procedural fairness and right process, (ii) the impacts of algorithms on administrative decisions are to be assessed, and harmful outcomes have to be reduced when found, (iii) data and information on the use of automated decision-making systems in federal institutions are made available to the public.

Unlike the legislative experiences promoted in other international settings, the Canadian directive has a more limited scope of application. Specifically, the directive is not a generally applicable rule that governs all AI systems, automated decision-making processes, or related systems across Canada. Instead, the scope of the DADM is limited to a narrow class of systems and activities within the Canadian federal government. In this respect, Section 5 provides that the directive applies:

- only to systems that provide external services as defined in the Service and Digital Policy.
- to any system, tool, or statistical model used to recommend or make an administrative decision about a customer.
- only to systems in production; automated decision-making systems operating in test environments are excluded.
- to any automated decision-making system developed or acquired after 1 April 2020.
- It does not apply to any national security system.

Therefore, the scope and application of the directive are subject to several important exceptions and limitations: only the systems of the federal government and federal agencies are regulated, but the use of any technology, not just those related to AI, is taken into account.⁴ However, the rules do not apply to private sector AI systems and systems used by provincial governments, municipalities, or provincial agencies, such as police services, children's services, and security services. Indeed, although the wording "*to any statistical system, tool or model used to recommend or make an administrative decision about a client*" may seem broad, Prof. Teresa Scassa pointed out that "*the Canadian directive focuses on decision-making [...] It is important to keep in mind that there may be many more choices/actions that do not formally qualify as decisions and that may impact on the lives of individuals or communities. These fall outside the directive and remain without specific governance.*"⁵

The DADM also features a 'risk-based' governance model like the European "Artificial Intelligence Act". In particular, four different levels of risk are established based on the impact that an automated decision will have on the rights of individuals or communities, the health or well-being of individuals or communities, the economic interests of individuals, entities, or communities, and the sustainability of an ecosystem:

1. **Level I:** the decision is likely to have little or no impact (impacts that are reversible and brief)
2. **Level II:** the decision is likely to have moderate impacts (impacts that are likely to be reversible and short term)
3. **Level III:** The decision is likely to have high impacts (impacts that may be difficult to reverse and are continuous).

⁴Statistics Canada, Responsible use of automated decision systems in the federal government - <https://www.statcan.gc.ca/en/data-science/network/automated-systems>

⁵Teresa Scassa, Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada's Directive on Automated Decision-Making (October 30, 2020). Forthcoming, (2021) 54:1 University of British Columbia Law Review [Scassa], online: <https://ssrn.com/abstract=3722192> at 6-7.

4. **Level IV:** The decision is likely to have very high impacts (impacts that are irreversible and perpetual).

Depending on the level of impact of the system, the directive may impose additional requirements, such as: peer reviews performed by qualified experts, involvement of people in decision-making processes, additional documentation on the design and functionality of the automated decision-making system, contingency plans, and backup systems in case the automated decision-making system becomes unavailable. The approval of these acts is the task of the Deputy Chief or the entire Treasury Council.

The impact of a given automated decision is calculated through the analysis of the **Algorithmic Impact Assessment**⁶, a supporting tool of the directive. The AIA is a crucial component of the Canadian strategy. Its operation is based on asking individuals or organisations considering a ADM system to answer approximately 60 questions designed to assess the appropriate level of risk for a system. The questions cover project details, the impact of a system, and proposed mitigation measures. Once the answers to these questions are incorporated into the AIA, a report is produced indicating the level of impact of the proposed systems and the associated requirements for a review. Finally, the final version of the AIA on the analysed system must then be published on the Government of Canada or Open Government websites (official government database).

Comments on the Canadian Strategy

The explicit objectives of the DADM are *"to ensure that automated decision-making systems are deployed in ways that reduce risk to Canadians and federal institutions, and lead to more efficient, accurate, consistent and interpretable decisions."* The first and perhaps most notable strength of the directive is its comprehensiveness. Indeed, it addresses an impressive range of issues, including:

- Basic requirements for the federal government's automated decision-making systems.
- Solid protections for the transparency of automated decision-making
- a mandatory register
- a detailed and considered risk assessment process
- a corrective regime
- a commitment to procedural fairness
- an oversight regime.

Among the strategy's strengths, we can highlight its **focus on automated government decision-making**. Unlike the European proposal contained in the Artificial Intelligence Act (2021), Canada's DADM is not a generally applicable rule or regulation that aims to regulate government and private sector AI and ADM simultaneously. Accordingly, **the directive addresses issues and concerns related to automated decision-making ADM systems applied specifically to the public sector and regulates and manages the operations of most ministries and public agencies in Canada**. The Directive also aims to favour the protection of individual rights instead of promoting the development or innovation of the AI market, a preference demonstrated by the Directive's extensive commitments to procedural fairness and comprehensive risk assessment.

On the other hand, the directive has received criticism concerning what has been deemed its short comings. As already pointed out, the DADM has a minimal scope, aiming to regulate a specific range of automated decision-making systems of the federal government. This means that other types of AI and algorithmic systems are beyond their scope. Moreover, the directive does not explicitly identify prohibited or 'high risk' systems, as is the case, for instance, in the proposed European law.

⁶Open Government - <https://open.canada.ca/data/en/dataset/748a97fb-6714-41ef-9fb8-637a0b8e0da1>

In general, one can conclude that the Canadian strategy for an ethical vision of AI is currently among the most the Canadian strategy for an ethical vision of AI is currently among the most valuable and incisive, even if applied to the restricted sphere of governmental activities.

UNITED STATES

National Artificial Intelligence Initiative Act

US legislation on digital ethics

Reference is made to two main acts:

- a) US Guidance for Regulation of AI Principles - 2020**
- b) National AI Initiative Act – 2021**

The basis of legislation

There has been more hesitation in the US than in the EU to introduce legal restrictions on AI, fearing that these might hamper innovation. In October 2016, the *White House Office of Science and Technology Policy (OSTP)* published the first US reports explicitly focused on AI: the '*Preparing for the Future of Artificial Intelligence*'⁷ and the '*National Artificial Intelligence Research and Development Strategic Plan*'.⁸ These reports defined the US government's role in AI development as one of innovation facilitator and minimalist regulator, applying a strategy that was judged as inadequate by many experts. Subsequently, during the Trump administration, government agencies were dissuaded from introducing new regulatory measures and continued with the approach to AI outlined in the 2016 documents. The government also formally stated that it had no intention of developing a national plan on AI; however, in 2019, this position was revised, and the Trump administration signed the '*American AI Initiative*'.⁹ With this executive order, the government underlined the US desire to maintain its leadership position in AI. Five basic principles are outlined in the document:

1. Driving technological advances in AI to promote scientific discovery, economic competitiveness, and national security.
2. Develop appropriate technical standards and reduce barriers to AI's safe testing and deployment.
3. Train workers with the right skills to prepare them for future jobs.
4. Promote public confidence in AI technologies and protect civil liberties, privacy, and American values.
5. Promote an international environment that supports research and opens markets for America's AI industries, protecting U.S. technological advantage and AI technologies from competitors.

The US Guide to the Regulation of AI Principles - 2020

Subsequent documents have drawn on these five principles but in a relatively limited area of governance. An important step forward was then taken in November 2020, with the publication by

⁷OSTP, *Preparing for the future of Artificial Intelligence*, 2016 -

https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

⁸OSTP, *National Artificial Intelligence Research and Development Strategic Plan*, 2016 -

https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf

⁹Executive Office of the President, *American AI Initiative*, 2019 -

<https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence>

the White House of the '*US Guidance for Regulation of AI Principles*.'¹⁰ This document is a binding guide for government agencies on the regulation of AI to be applied to the private sector, focusing on three main themes: (i) limiting over-regulation; (ii) ensuring public involvement; and (iii) promoting reliable, fair, transparent and secure AI. The guide also specifies ten fundamental principles for AI management:

- a) **Public trust in AI** (government must promote reliable, robust, and trustworthy AI applications)
- b) **Public participation** (the public must have the opportunity to participate in all stages of the regulatory process)
- c) **Scientific integrity and quality of information** (policy decisions must be based on science)
- d) **Risk assessment and management** (government agencies must decide which risks are unacceptable)
- e) **Benefits and costs** (government agencies must select approaches that maximise net benefits)
- f) **Flexibility** (government agencies must pursue a flexible and technology-neutral approach)
- g) **Fairness and non-discrimination** (government agencies must ensure that AI systems do not unlawfully discriminate)
- h) **Disclosure and transparency** (specific transparency measures needed to gain public confidence)
- i) **Safety and security** (government agencies must promote AI systems that are safe, secure, and operate as intended)
- j) **Coordination among government agencies** (cooperation and coordination among government agencies are necessary to ensure consistent policies)

The National AI Initiative Act - 2021

Recently, *The National AI Initiative Act*¹¹ of 2020, passed in January 2021, codified the US vision of AI into binding regulations. The goals of the Act largely reflect those already contained in the “American AI Initiative”, which focuses on American AI leadership in the research and development of reliable AI systems, as well as in preparing the workforce for the potential impact of AI on work processes and coordination between the military and civilian sectors. The definition of AI for the US system is provided in Title I, Section 3 of the Act:

- **Artificial Intelligence:** the term “artificial intelligence” means a machine-based system that can, for a given set by the men defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs i) to perceive real and virtual environments; ii) abstract such perceptions into models through analysis carried out in an automated manner; and use model inference to formulate options for information or action.

The Act provides for the establishment of several bodies to provide guidance at the federal level:

- **‘National Artificial Intelligence Initiative Office’** (Section 102) operates under the direction of the White House's Office of Science and Technology Policy-OSTP to drive global US leadership in the development and use of reliable AI systems and prepare the nation's workforce for the 'integration of AI into all sectors of the economy. Therefore, the

¹⁰Executive Office of the President. (2020). Guidance for regulation of Artificial Intelligence applications. <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>

¹¹US Congress,National AI Initiative Act, 2021 - <https://www.congress.gov/bill/116th-congress/house-bill/6216/text#toc-H06C2598E670D416096729065FCC48CF9>

body's mission is to serve as a point of contact for federal AI activities for executive branch departments and agencies and other public and private entities that may be involved in the initiative.

- **‘Interagency AI Committee’** (Section 103), co-chaired by the White House Director of OSTP and representatives from the Departments of Commerce and Energy and the National Science Foundation (NSF) on an annual rotating basis, to manage interagency activities and planning.
- **‘National AI Advisory Committee’** (Section 104), established by the Secretary of Commerce, works in consultation with the Director of OSTP, the Director of National Intelligence (DNI), and the Secretaries of Defense, Energy, and State, to assess whether the United States is achieving its goals.

The bill also allowed the appropriation of \$4.8 billion through the year 2025 to fund *National Science Foundation (NSF)* programmes and support AI research and the training of a skilled AI workforce. The NSF has also been ordered to contract with the *National Research Council - National Academies of Sciences, Engineering, and Medicine* to study the current and future impact of AI on the US workforce across all sectors. In this regard, within two years of the law's approval, the NSF is to provide a report that includes findings and recommendations to be submitted to Congress later. The NSF has also been authorised to establish a network of research institutes focusing on cross-cutting challenges for AI systems, such as reliability, or focusing on a particular economic or social sector, such as health care, education, and manufacturing. To mitigate the potential social risks involved in conducting the research, these institutes will also have to draw up an ethical statement for all AI research proposals. Finally, thanks to the law, the *National Institute of Standards and Technology (NIST)*, under the control of the Department of Commerce, will have to include in its objectives the creation of a set of guidelines for AI, supporting the development of a framework for mitigating risk and promoting reliable AI systems.¹².

Comments on the US strategy

Overall, the “National AI Initiative Act” aims to accelerate and coordinate federal investment in AI by facilitating new public-private partnerships in AI research, standard setting and education, in order to guarantee the United States a leading global positioning in the development and use of reliable AI systems. On the ethical side, the legislation establishes the development of standards at NIST and requires a framework for managing the risks associated with AI systems. It also supports research at NSF in a wide variety of AI-related research areas, both to improve AI systems and to use those systems to advance other areas of science.

Basically, the US approach is characterised by promoting innovation to maintain global US leadership in AI while limiting over regulation as much as possible. The National AI Initiative Act certainly marks a change concerning previous governance choices. Nevertheless, the United States has preferred to focus on enhancing the **positive freedoms of individuals and businesses to benefit from AI rather than emphasising the restrictions associated with a negative idea of freedom and thus promoting regulations aimed primarily at safeguarding and protecting individual freedoms**. The US approach is adopted to foster the AI economy as much as possible, despite many experts have repeatedly criticised the assumption that limiting regulation is the best way to ensure innovation.

Steps have been taken to materialise an ethical vision of AI in the US. For example, the government has committed to doubling federal AI research and development funding for non-defence projects in

¹²DLA Piper, Landmark artificial intelligence legislation advances toward becoming law, 2020 - <https://www.dlapiper.com/en/us/insights/publications/2020/12/landmark-artificial-intelligence-legislation-advances-toward-becoming-law/>

2021¹³. Through education and training initiatives, efforts to prepare the US for societal and labor market change have also increased. For example, proposals have been made to improve STEM education and boost research through a range of scholarships and training programmes. However, experts doubt whether even these programmes can address the social changes that AI could cause, particularly among the less educated parts of the population. Socio-economic differences between the different groups that make up the US population have been overlooked. These differences could mean that the social disruptions associated with the spread of AI will be more damaging to marginalised groups that have not had access to education and re-training initiatives.

Federal efforts to ensure that AI is ethical, safe, and trustworthy have been relatively few, limited almost exclusively to the principles listed in the US Guidance for Regulation of AI Principles and the invitation to refer to them. However, even if binding, these principles are only very general guidelines. In conclusion, from the perspective of ethics, according to numerous experts the US vision is insufficient given the many ethical challenges associated with the unlimited development and dissemination of AI technologies. The White House's emphasis on avoiding over-regulation is considered a disincentive for federal agencies to introduce strict regulations. In the past, for instance, some White House officials have criticised the states that have considered banning facial recognition technology. As reported in the research by Prof. Roberts et al. (2021): *'While self-regulation by industry can mitigate some potential ethical harms, the lack of specific regulatory measures and oversight can lead to practices such as **ethics-washing** (introducing superficial measures), **ethics shopping** (choosing ethical frameworks that justify actions after the fact) and **ethics lobbying** (exploiting digital ethics to delay regulatory standards). In practice, inadequate regulatory measures in the private sector have facilitated numerous harms, such as those arising from bias in facial and emotion recognition technologies that lead to discrimination. To see the potential for broader damages that this approach can cause, one need only look at previous failures of self-regulation without adequate oversight in other high-risk areas, such as the self-regulation of credit agencies that led to the 2007-2008 financial crisis and the airline industry before the Boeing 737 crashes'*¹⁴.

RUSSIAN FEDERATION

Code of Ethics on Artificial Intelligence

Introduction

After a long period of in-depth study and preparation involving in particular hundreds of scientists and experts from the public and private sectors, the Code of Ethics of the Russian Federation was officially presented on 26 October 2021 and came into force as binding law on 1 November 2021

The basis of Russian legislation

In September 2017, Russian President Vladimir Putin declared that the 'masters of AI will rule the world', thus announcing Russia's official entry into the 'global race' to develop artificial intelligence technologies.¹⁵ As early as 2016, Russian governmental organisations and companies had begun to

¹³Office of Science and Technology Policy. (2020). President Trump's FY 2021 budget commits to double investments in key industries of the future – The White House. <https://trumpwhitehouse.archives.gov/briefings-statements/president-trumps-fy-2021-budget-commits-double-investments-key-industries-future/>

¹⁴Roberts, H., Cows, J., Hine, E. et al. Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US. *Sci Eng Ethics* 27, 68 (2021) - <https://doi.org/10.1007/s11948-021-00340-7>

¹⁵Corriere della Sera (2017) -https://www.corriere.it/tecnologia/economia-digitale/17_settembre_04/putin-sull-intelligenza-artificiale-chi-sviluppa-migliore-governa-mondo-musk-rilancia-l-allarme-c2a46c9c-916f-11e7-8332-148b1c29464d.shtml

embrace the use of AI, such as Sberbank, a state-owned bank that created a venture capital fund focused on investing in startups in financial technology, big data, and artificial intelligence.¹⁶ The first major Russian government proposal on AI is contained in a ten-point statement by the Russian Ministry of Defence, presented in March 2018¹⁷. The proposal was the result of coordinated action between the Ministry of Defence, the Ministry of Education and Science, and the Russian Academy of Sciences (RAS), promoted to analyse the state of AI in Russia and to unite the main Russian educational, industrial and governmental organisations on the problems opened up by the use of AI technologies. In particular, the document responded to the need to promote a new Ministry of Defence research campus in the Black Sea area to provide the armed forces with innovative AI solutions. This ten-point proposal for AI development thus responded to specific military needs and identified actions to enable government bodies to take the lead in developing AI technologies in Russia. The ten points are as follows¹⁸:

1. Form an AI and Big Data consortium to combine the primary Russian efforts in AI implementation.
2. Intensify efforts to establish a fund for analytical algorithms and programs to provide expertise on automated systems.
3. Create a state system for AI training and education.
4. Establish an AI laboratory at ERA Technopolis in Anapa for AI research.
5. Establish a national centre for artificial intelligence to develop promising AI projects and implement AI solutions.
6. Monitor global AI development.
7. Hold AI "wargames" (military exercises carried out to test or improve tactical skills) organised by the Ministry of Defence.
8. Monitor IA compliance.
9. Discuss IA proposals in national military forums.
10. Hold an annual IA conference.

Then, also in 2018, President Putin issued a series of decrees outlining Russia's national development goals until 2024. These goals included increasing Russian life expectancy to 78 years, halving poverty, and introducing digital technologies into the economy and social sphere.¹⁹ Among the national projects envisaged by these decrees was the 'National Digital Economy Project.' In this project, the government mapped out the future of digital in the country through two parallel AI development initiatives. The first initiative was the 'Digital Technologies Federal Project', a digital project that aimed at developing seven end-to-end digital technologies. In addition to AI, it also focused on wireless communications (5G), robotics, virtual reality, and new manufacturing technologies. The second initiative, which began with the 'National Strategy for the Development of AI' and culminated in the 'AI Federal Project,' focused exclusively on AI.

In 2019, the government also commissioned state-owned companies to draw up implementation and financing plans called 'roadmaps' to develop new technologies. In this way, the **'Roadmap for Artificial Intelligence'** was approved, outlining the methods for advancing Russia's digital technology sector, along with details of the funding allocated for this purpose. In the same year,

¹⁶Stephanie Petrella, Chris Miller, and Benjamin Cooper, Russia's Artificial Intelligence Strategy: The Role of State-Owned Firms, November 2020 - <https://sites.tufts.edu/hitachi/files/2021/02/1-s2.0-S0030438720300648-main.pdf>

¹⁷"Conference "Artificial Intelligence: Problems and Solutions – 2018", Ministry of Defense of the Russian Federation, 2018 - <http://mil.ru/conferences/is-intellekt.htm>.

¹⁸Samuel Bendett, "Here's How the Russian Military Is Organizing to Develop AI," Defense One, July 20, 2018, <https://www.defenseone.com/ideas/2018/07/russian-militarys-ai-development-roadmap/149900/>.

¹⁹"The President signed Executive Order On National Goals and Strategic Objectives of the Russian Federation through to 2024," President of Russia, May 7, 2018 - <http://en.kremlin.ru/events/president/news/57425>

President Putin mandated the government to create a *National Strategy for AI development until 2030*.²⁰

Finally, in December 2020, during a virtual conference on AI, President Putin proposed the development of a code of ethics for artificial intelligence. The code goes beyond the need to introduce simple legislative or administrative restrictions and emphasising the need to develop a moral code as soon as possible to act as a guiding tool within the AI system.

The Code of Ethics for Artificial Intelligence - 2021

On 26 October 2021, the Russian government, the AI Alliance and various other organisations and companies signed the '*Code of Ethics for Artificial Intelligence*'.²¹ The signing took place in Moscow during the first international forum, 'Ethics of Artificial Intelligence: The Beginning of Trust.' The document refers to all the negative scenarios of new technologies feared by leading system developers.

The code is part of the *Artificial Intelligence Federal Project* and the *Strategy for the Development of the Information Society for 2017-2030*.²² The code establishes the general ethical principles and standards of conduct to be followed by participants in the system of relationships and activities promoted in the field of artificial intelligence (or "AI Actors," as they are defined in the document), as well as the mechanisms for implementing the provisions of the code. The provisions of the code apply to all ethical aspects of AI technologies relating to:

- **Creation**
- **Design**
- **Construction**
- **Piloting**
- **Implementation**

We should underline that the provisions of the code in the field of AI are addressed exclusively to civilian purposes, excluding the military sphere, and are intended to avoid the possible abuses and catastrophic scenarios that are occurring or could occur in the future through the use of artificial intelligence. The document is divided into two sections, including the areas of activity relative to the encouragement of the AI development, the awareness of ethics in the AI use, the identification of AI in communication with a person, and the security of information. The various paragraphs of the Sections then provide a catalogue of the negative impacts of AI on society and, concerning each possible effects, provide an ethical assessment to be applied in the preventive phase. Also, in this case, the code presents a "risk-based" approach, as is specified in Section II, clause 2.1, which states:

Risk-based approach

The level of attention to ethical issues in AI and the nature of the relevant actions of AI Actors should be proportional to the assessment of the level of risk posed by specific technologies and AI systems to the interests of individuals and society. Risk-level assessment must take into account both the known and possible risks; in this case, the level of probability of threats should be taken into account as well as their possible scale in the short and long term. In the field of AI

²⁰“Rosatom summed up the results of tenders for the development of ‘roadmaps’ for end-to-end technologies”, D-Russia.ru, April 1, 2019 - <https://d-russia.ru/rosatom-podvelitogi-konkursov-na-razrabotku-dorozhnyh-kart-po-skvoznym-tehnologiyam.html>.

²¹Artificial Intelligence Code of Ethics - <https://a-ai.ru/wp-content/uploads/2021/10/Code-of-Ethics.pdf>

²²TASS Russian News Agency, First code of ethics of artificial intelligence signed in Russia, 26 OCT 2021 - <https://tass.com/economy/1354187>

development, making decisions that are significant to society and the state should be accompanied by scientifically verified and interdisciplinary forecasting of socio-economic consequences and risks, as well as by the examination of possible changes in the value and cultural paradigm of the development of society, while taking into account national priorities. In pursuance of this code, the development and use of an AI system risk assessment methodology is recommended.

The *AI Alliance* drafted the code together with the *Analytical Centre of the Russian Government* and the *Ministry of Economic Development*. It was signed by representatives of major economic players such as *Sberbank, Gazprom Neft, Yandex VK, MTS, Russian Direct Investment Fund, the Skolkovo Research Cluster, Rostelecom, Rosatom, InfoWatch, the Cianreal estate platform*. Adherence to the code is voluntary, as set out in Section II, paragraph 2, clause 2.1.

Paragraph 1 of Section I opens with clause 1.1. It establishes that in the development of AI technologies, a person's rights and freedoms must be considered of the highest value. The use of such technologies must promote and not hinder the realisation of all potential human capabilities. Specifically, it establishes that AI developers should be guided by a humanistic approach, meaning that their models and related instrumentation should not operate at the expense of people's rights and interests, nor discriminate against or negatively affect their cognitive abilities, socialisation and morality. The subsequent clauses of paragraph 1 then set out general ethical principles, including:

- **1.2. Respect for human autonomy and freedom of will:** AI developers are required “to anticipate possible negative effects on the development of human cognitive abilities and prevent the development of AI systems that intentionally cause such effects”, with the aim to preserve the autonomy and free will of a human's decision-making ability, the right to choose, and, in general, the intellectual abilities of an individual as an independent value and a system-forming factor of modern civilization.
- **1.3. Compliance with the law:** AI Actors must know and comply with the provisions of the legislation of the Russian Federation in all areas of their activities and at all stages of the creation, development and use of AI technologies, including in matters of the legal responsibility of AI Actors.
- **1.4. Non-discrimination:** To ensure fairness and non-discrimination, AI Actors should take measures to verify that the algorithms, data sets and processing methods for machine learning that are used to group and/or classify data concerning individuals or groups do not intentionally discriminate. AI Actors are encouraged to create and apply methods and software solutions that identify and prevent discrimination based on race, nationality, gender, political views, religious beliefs, age, social and economic status, or information about private life. (At the same time, cannot be considered as discrimination rules, those which are explicitly declared by an AI Actor for functioning or the application of AI systems with reference to the different groups of users, taking into account the need to proceed sometime with segmentation processes).

An important aspect to consider is contained in clause 2.9 of Section I, entitled "Recursive control of self-improvement of strong AI." In this clause, AI Actors are encouraged to cooperate in identifying and verifying information on the methods and forms of creating so-called "strong" AI and preventing the possible threats they represent. In particular, recursive self-improvement is a process in which a programme improves itself not only in terms of what it can do, but also in terms according to which its algorithms come to 'decide' what to do. This type of AI system would therefore be able to collect information, predict, communicate in natural language; it would also have the properties of will and character. For the moment, this type of technology is only possible in theory. Still, since it is foreseeable that one day strong AI systems may have the capacity for

unlimited recursive self-improvement, it is of main importance that Russia has already included them in the provisions of the code, taking into account possible negative scenarios.²³

In paragraph 5 of Section I, on the other hand, the code places the interests of AI development above the competition, i.e., it encourages developers to cooperate and jointly improve skills. For this reason it is provided that information on AI should be transparent and accessible so that both the industry and users have a clear understanding of the level of technological development and the associated risks.

Section II, finally, contains the provisions concerning the essential operation of the code, some definitions and the mechanisms for linking and implementing the provisions, establishing the creation of the figures of ***Ethics Commissioners*** and a ***Code Implementation Commission***, which will include scientists, business representatives and employees of government agencies. Indeed, despite the consultative nature of the Code, every organisation that decides to adhere to its provisions will be obliged to appoint IA Ethics Commissioners who will monitor compliance.

3 - Comments on the strategy of the Russian Federation

The Russian Code certainly represents a step forward in the field of AI ethics. Among the main points of the new regulatory model is the **voluntary compliance** of companies with the Code, a fact that if, on the one hand, it limits the area of application of the law, on the other hand it leaves companies with full responsibility for the choice about the best way to acquire credibility and trust in the market and in society, all elements objectively linked to adhering to the Code of Ethics. Another important aspect concerns the creation of the **Ethics Commissioners** who will operate in national and regional institutions and within companies. A new important body in particular for the confrontation with the social partners and for the strengthening of industrial relations in this matter

Experts pointed out some criticalities concerning the document and its regulatory capacity. For instance, *Andrey Filchenkov*, head of the Machine Learning Lab at ITMO University, pointed out that adherence to the code is voluntary and does not apply to military developments. This means that by violating the recommendations of the Code, companies only risk reputational losses in the worst case scenario.

In addition, *Andrei Krichevsky*, President of the IP Chain Association, pointed out that the code does not regulate the relatively new issue for the industry relating to the intellectual property of what an AI system can create. Today, a significant proportion of creative content is made using AI, and the lack of clearly defined rules on ownership creates several difficulties in the industry. Krichevsky believes that these problems can be avoided if exclusive copyrights are granted not only to natural persons but also to legal entities, following the example of how record labels manage exclusive rights on phonograms²⁴.

The Text of the Artificial Intelligence Code of Ethics

The Code of Ethics for Artificial Intelligence (hereinafter referred to as the Code) establishes the general ethical principles and standards of conduct that should be followed by participants to the relationships to the field of artificial intelligence (hereinafter referred to as AI Actors) in their activities, as well as the mechanisms for the implementation of the provisions of this Code. The Code applies to relationships related to the ethical aspects of the creation (design, construction,

²³Forbes, Tutte le paure del mondo: cosa temono i creatori del Codice Etico dell'Intelligenza Artificiale, 1° novembre 2021 - <https://www.forbes.ru/tekhnologii/444503-vse-strahi-mira-cego-boatsa-sozdateli-kodeksa-etiki-iskusstvennogo-intellekta>

²⁴Agenzia Sputnik, Non nuocere: gli esperti hanno raccontato cosa cambierà dopo la comparsa del Codice Etico per l'Intelligenza Artificiale, 1° novembre 2021 - <https://spbdnevnik.ru/news/2021-11-01/ne-navredi-eksperty-rasskazali-chto-izmenitsya-posle-poyavleniya-kodeksa-etiki-iskusstvennogo-intellekta>

piloting), implementation and use of AI technologies at all stages that are currently not regulated by the legislation of the Russian Federation and/or by acts of technical regulation. The recommendations of this Code are designed for artificial intelligence systems (hereinafter referred to as AI Systems-AIS) used exclusively for civil (not military) purposes. The provisions of the Code can be expanded and/or specified for individual groups of AI Actors with sectoral or local documents on ethics in the field of AI, considering the development of technologies, the specifics of the tasks to be solved, the class and purpose of the AIS and the level of possible risks, as well as the specific context and environment in which the AIS are being used.

SECTION I PRINCIPLES OF ETHICS AND RULES OF CONDUCT

1. THE MAIN PRIORITY OF THE DEVELOPMENT OF AI TECHNOLOGIES IS PROTECTING THE INTERESTS AND RIGHTS OF HUMAN BEINGS COLLECTIVELY AND AS INDIVIDUALS

1.1. Human-centered and humanistic approach. In the development of AI technologies, the rights and freedoms of the individual should be given the greatest value. AI technologies developed by AI Actors should promote or not hinder the realization of humans' capabilities to achieve harmony in social, economic and spiritual spheres, as well as in the highest self-fulfillment of human beings. They should take into account key values such as the preservation and development of human cognitive abilities and creative potential; the preservation of moral, spiritual and cultural values; the promotion of cultural and linguistic diversity and identity; and the preservation of traditions and the foundations of nations, peoples and ethnic and social groups. A human-centered and humanistic approach is the basic ethical principle and central criterion for assessing the ethical behavior of AI Actors, which are listed in the section 2 of this Code.

1.2. Respect for human autonomy and freedom of will. AI Actors should take all necessary measures to preserve the autonomy and free will of a human's decision-making ability, the right to choose, and, in general, the intellectual abilities of a human as an intrinsic value and a system-forming factor of modern civilization. AI Actors should, during AIS creation, assess the possible negative consequences for the development of human cognitive abilities and prevent the development of AIS that intentionally cause such consequences.

1.3. Compliance with the law. AI Actors must know and comply with the provisions of the legislation of the Russian Federation in all areas of their activities and at all stages of the creation, development and use of AI technologies, including in matters of the legal responsibility of AI Actors.

1.4. Non-discrimination. To ensure fairness and non-discrimination, AI Actors should take measures to verify that the algorithms, datasets and processing methods for machine learning that are used to group and/or classify data concerning individuals or groups do not intentionally discriminate. AI Actors are encouraged to create and apply methods and software solutions that identify and prevent discrimination based on race, nationality, gender, political views, religious beliefs, age, social and economic status, or information about private life. (At the same time, according to the new rules, it cannot be considered as a discrimination act that explicitly declared and promoted by an AI Actor for the functioning or the application of AIS to the different groups of users, when such elements are taken into account for the need/choice of the segmentation processes)

1.5. Assessment of risks and humanitarian impact. AI Actors are encouraged to assess the potential risks of using an AIS, including the social consequences for individuals, society and the state, as well as the humanitarian impact of the AIS on human rights and freedoms at different stages, including during the formation and use of datasets. AI Actors should also carry out long-term monitoring of the manifestations of such risks and take into account the complexity of the behavior of AIS during risk assessment, including the interconnection and the interdependence of

processes in the AIS's life cycle. For critical applications of the AIS, in special cases, the AI Actor is encouraged to conduct of the risk assessment with the support and the involvement of a neutral third party or authorized official body, but without prejudice to the performance and information security of the AIS as well as to the protection of the intellectual property and trade secrets of the developer.

2. NEED FOR CONSCIOUS RESPONSIBILITY WHEN CREATING AND USING AI

2.1. Risk-based approach. The level of attention to ethical issues in AI and the nature of the relevant actions of AI Actors should be proportional to the assessment of the level of risk posed by specific technologies and AISs and the interests of individuals and society. Risk-level assessment must take into account both the known and possible risks; in this case, the level of probability of threats should be taken into account as well as their possible scale in the short and long term. In the field of AI development, making decisions that are significant to society and the state should be accompanied by scientifically verified and interdisciplinary forecasting of socio-economic consequences and risks, as well as by the examination of possible changes in the value and cultural paradigm of the development of society, while taking into account national priorities. In pursuance of this Code, the development and use of an AIS risk assessment methodology is recommended.

2.2. Responsible attitude. AI Actors should be accountable for the impact of AI systems on society and its citizens at every stage of its life cycle, including the protection of privacy, the ethical, safe and responsible use of personal data, the nature, the degree and substance of the damage that can derive from the use of AI technology and systems, as well as from the selection and use of hardware and software used in the various stages of implementation of the AI Systems. At the same time, the responsibility of the AI Actors must correspond to the nature, degree and amount of damage that may occur as a result of the use of technologies and AIS; it is also necessary to take into account the role of the AI Actor in the life-cycle of AIS, as well as the degree of possible and real impact of a particular AI Actor on causing damage, as well as its size.

2.3. Precautions. When the activities of AI Actors can lead to morally unacceptable consequences for individuals and society, the occurrence of which can reasonably be traced back to the corresponding AI Actor, the last should take reasonably measures to prevent or limit the occurrence of such consequences. To assess the category of "moral unacceptable consequences" and discuss possible measures to prevent them, Actors can use the provisions of the Code, including the mechanisms specified in Section 2.

2.4. No harm. AI Actors should not allow use of AI technologies for the purpose of causing harm to human life, the environment and/or the health or property of citizens and legal entities. Any application of an AIS capable intentionally of causing harm to the environment, human life or health or the property of citizens and legal entities during any stage, including design, development, testing, implementation or operation, is unacceptable.

2.5. Identification of AI in communication with a human. AI Actors are encouraged to ensure that users are informed of their interactions with the AIS when it affects their rights and critical areas of their lives and to ensure that such interactions can be terminated at the request of the user.

2.6. Data security. AI Actors must: comply with the legislation of the Russian Federation in the field of personal data and secrets protected by law when using an AIS. Furthermore; they must ensure the protection of personal data processed by an AIS or AI Actors with the aim to develop and improve the AIS; develop and implement innovative methods of controlling unauthorized access by third parties to personal data; to use high-quality and representative datasets from reliable sources and obtained without violating the law.

2.7. Information security. AI Actors should provide the maximum possible protection against unauthorized interference in the work of the AI by third parties by introducing adequate information security technologies, including the use of internal mechanisms for protecting the AIS from unauthorized interventions and informing users and developers about such interventions. They must also inform users about the rules regarding information security when using the AIS.

2.8. Voluntary certification and Code compliance. AI Actors can implement voluntary certification for the compliance of the developed AI technologies with the standards established by the legislation of the Russian Federation and this Code. AI Actors can create voluntary certification and AIS labeling systems that indicate that these systems have passed voluntary certification procedures and confirm quality standards.

2.9. Control of the recursive self-improvement of AISs. AI Actors are encouraged to collaborate in the identification and verification of methods and forms of creating universal ("strong") AIS and the prevention of the possible threats that AIS carry. The use of "strong" AI technologies should be under the control of the state.

3. HUMANS ARE ALWAYS RESPONSIBLE FOR THE CONSEQUENCES OF THE APPLICATION OF AN AIS

3.1. Supervision. AI Actors should provide comprehensive human supervision of any AIS, whose extent and manner is depending from the purpose of the single AIS, including, for example, recording significant human decisions at all stages of the AIS life cycle or making provisions for the registration of the work of the AIS. They should also ensure the transparency of AIS use, including the possibility of cancellation by a person and (or) the prevention of making socially and legally significant decisions and actions by the AIS at any stage in its life cycle, where reasonably applicable.

3.2. Responsibility. AI Actors should not allow the transfer of rights of responsible moral choice to the AIS or delegate responsibility for the consequences of the AIS's decision-making. A person (an individual or legal entity recognized as the subject of responsibility in accordance with the legislation in force of the Russian Federation) must always be responsible for the consequences of the work of such systems. AI Actors are encouraged to take all measures to determine the responsibilities of specific participants in the life cycle of the AIS, taking into account each participant's role and the specifics of each stage.

4. AI TECHNOLOGIES SHOULD BE APPLIED AND IMPLEMENTED WHERE IT WILL BENEFIT PEOPLE

4.1. Application of AIS in accordance with its intended purpose. AI Actors must use AIS in accordance with the stated purpose, in the prescribed subject area and for solving the prescribed problems.

4.2. Stimulating the development of AI. AI Actors should encourage and incentivize the design, implementation, and development of safe and ethical AI technologies, taking into account national priorities.

5. INTERESTS OF DEVELOPING AI TECHNOLOGIES ABOVE THE INTERESTS OF COMPETITION

5.1. Correctness of AIS comparisons. To maintain the fair competition and effective cooperation of developers, AI Actors should use the most reliable and comparable information about the capabilities of AISs in relation to a task and ensure the uniformity of the measurement methodologies.

5.2. Development of competencies. AI Actors are encouraged to follow practices adopted by the professional community, to maintain the proper level of professional competence necessary for safe and effective work with AIS and to promote the improvement of the professional competence of workers in the field of AI, including within the framework of programs and educational disciplines on AI ethics.

5.3. Collaboration of developers. AI Actors are encouraged to develop cooperation within the AI Actor community, particularly between developers, including by informing each other of the identification of critical vulnerabilities in order to prevent their wide distribution. They should also

make efforts to improve the quality and availability of resources in the field of AIS development. Moreover also for:

- *to increase the availability of data (including labeled data);
- *to ensure the compatibility of the developed AIS where applicable;
- *to create conditions for the formation of a national school for the development of AI technologies that includes publicly available national repositories of libraries and network models, available national development tools, open national frameworks, etc.;
- *to share information on the best practices in the development of AI technologies;
- *to organize and hold conferences, hackathons and public competitions, as well as high-school and student Olympiads;
- *to increase the availability of knowledge and encourage the use of open knowledge databases;
- °to create conditions for attracting investments in the development of AI technologies from Russian private investors, business angels, venture funds and private equity funds while stimulating scientific and educational activities in the field of AI by participating in the projects and activities of leading Russian research centers and educational organizations.

6. IMPORTANCE OF MAXIMUM TRANSPARENCY AND RELIABILITY IN INFORMATION ON THE LEVEL OF DEVELOPMENT, CAPABILITIES AND RISKS OF AI TECHNOLOGIES

6.1. Credibility of information about AIS. AI Actors are encouraged to provide AIS users with credible information about such systems, the acceptable and most effective methods of using the AIS and about the harm, benefits, and existing limitations of their use.

6.2. Raising awareness of the ethics of AI application. AI Actors are encouraged to carry out activities aimed at increasing the level of trust and awareness of citizens who use AISs and society in general. This should include increasing awareness of the technologies being developed, the features of the ethical use of AISs and other provisions accompanying the development of AIS. This promotion could include the development of journal articles, the organization of scientific and public conferences and seminars, and the inclusion of rules of ethical behavior for users and operators in the rules of the AIS.

SECTION 2 APPLICATION OF THE CODE

1. FOUNDATION OF THE CODE ACTION

1.1. Legal basis of the Code. The Code takes into account the legislation of the Russian Federation, the Constitution of the Russian Federation and other regulatory legal acts and strategic planning documents. These include the *National Strategy for the Development of Artificial Intelligence*, the *National Security Strategy of the Russian Federation* and the *Concept for the Regulation of Artificial Intelligence and Robotics*. The Code also considers international treaties and agreements ratified by the Russian Federation applicable to issues ensuring the rights and freedoms of citizens in the context of the use of information technologies.

1.2. Terminology. Terms and definitions in this Code are fixed in accordance with applicable regulatory legal acts, strategic planning documents and technical regulation in the field of AI.

1.3. AI Actors. For the purposes of this Code, AI Actors are defined as persons, including foreign ones, participating in the life cycle of an AIS during its implementation in the territory of the Russian Federation or in relation to persons who are in the territory of the Russian Federation, including those involved in the provision of goods and services. Such persons include, but are not limited to, the following:

- *developers who create, train, or test AI models/systems and develop or implement such models/systems, software and/or hardware systems and take responsibility for their design;

- *customers (individuals or organizations) receiving a product; or a service;
- *data providers and persons involved in the formation of datasets for their use in AISs;
- *experts who measure and/or evaluate the parameters of the developed models/systems;
- *manufacturers engaged in the production of AIS;
- *AIS operators who legally own the relevant systems, use them for their intended purpose and directly implement the solution to the problems that arise from using AIS;
- *operators (individuals or organizations) carrying out the work of the AIS;
- *persons with a regulatory impact in the field of AI, including the developers of regulatory and technical documents, manuals, various regulations, requirements, and standards in the field of AI; and
- *other persons whose actions can affect the results of the actions of an AIS or persons who make decisions on the use of AIS.

2. MECHANISM OF ACCESSION AND IMPLEMENTATION OF THE CODE

2.1 Voluntary Accession. Joining the Code is voluntary. By joining the Code, AI Actors agree to follow its recommendations. Joining and following the provisions of this Code may be taken into account when providing support measures or other interactions with an AI Actor or between AI Actors.

2.2 Ethics officers and/or ethics commissions. To ensure the implementation of the provisions of this Code and the current legal norms when creating, applying and using an AIS, AI Actors appoint officers on AI ethics who are responsible for the implementation of the Code and who act as contacts for AI Actors on ethical issues involving AI. These officers can create collegial industry bodies in the form of internal ethics commissions in the field of AI to consider the most relevant or controversial issues in the field of AI ethics. AI Actors are encouraged to identify an AI ethics officer when ever possible upon accession to this Code or within two months from the date of accession to the Code.

2.3. Commission for the Implementation of the National Code in AI Ethics. In order to implement the Code, a commission for the implementation of the Code in the field of AI ethics (hereinafter referred to as the Commission) is established. The commission may have working bodies and groups consisting of representatives of the business community, science, government agencies and other stakeholders. The Commission considers the applications of AI Actors wishing to join the Code and follow its provisions; it also maintains a register of Code members. The activities of the Commission and the conduct of its secretariat are carried out by the *Alliance for Artificial Intelligence Association* with the participation of other interested organizations.

2.4. Register of Code participants. To accede to this Code, the AI Actor sends a corresponding application to the Commission. The register of AI Actors who have joined the Code is maintained on a public website/portal.

2.5. Development of methods and guidelines. For the implementation of the Code, it is recommended to develop methods, guidelines, checklists and other methodological materials to ensure the most effective observance of the provisions of the Code by the AI Actors

2.6. Code of Practice. For the timely exchange of best practices, the useful and safe application of AIS, built on the basic principles of this Code, increasing the transparency of developers' activities, and maintaining healthy competition in the AIS market, AI Actors may create a set of best and/or worst practices for solving emerging ethical issues in the AI life cycle, selected according to the criteria established by the professional community. Public access to this code of practice should be provided.

EUROPEAN UNION

The EU proposal: the Artificial Intelligence Act

Introduction

In 2016, the Commission published the first document on the EU management of AI. The EU initiative continued until the official presentation on 21 April 2021 of a specific organic proposal, the *Artificial Intelligence Act*, which started its legislative process to become a binding act of the Union. Therefore, in the case of the EU, we are dealing with a proposal, not a law in force. It is a proposal that stands out for its organic nature in tackling the complex problems opened up by Artificial Intelligence, starting with those relating to the ethical principles of reference of the development processes. Faced with the process of translating the proposal into law, it will be important to understand what possible changes will be made to the initial text and their reasons.

The foundations of the Artificial Intelligence Act

In May 2016, the European Union published its first document addressing the issue of EU-wide management of AI. The paper, entitled '*Civil Law Rules on Robotics*'²⁵ and drafted by the Legal Affairs Committee of the European Parliament (JURI), introduced a coordinated European approach through several possible 'hard' and 'soft' law measures to protect citizens from potential risks related to the use of AI. Since the 2016 JURI report, the attention on AI by EU policy makers has increased significantly. In April 2018, 24 European countries, together with Norway, signed the '*EU Declaration on Cooperation on Artificial Intelligence*'²⁶, with which they declared their intention to promote a common European response to the open issues of AI development and applications.

In the same year, the European Commission published the '*Communication on the European Approach to AI*,' which outlined a coordinated approach centred on three priorities:

1. Increase the EU's technological and industrial capacity throughout the economy, in both the private and public sectors.
2. Prepare for the changes brought about by AI by anticipating market changes, modernising education and training, and adapting social protection systems.
3. Ensure the existence of an appropriate legal and ethical framework consistent with EU values.

Then, in 2019, the European Commission defined a list of non-binding ethical guidelines for building reliable AI²⁷. Prepared by the Commission's "*High-Level Expert Group on AI*" (HLEG), composed of 52 independent experts, this document aimed to offer guidance on promoting and ensuring the development of ethical AI systems in the EU. In the document, the European Commission stated that while recognising the potential and concrete benefits to individuals and society, some AI applications could also have adverse effects, including complex products to predict, detect or measure. For this reason, the group of experts, authors of the guidelines, thought of laying the foundations for reliable AI through the identification of the following fundamental principles: i) Respect for human dignity, ii) Individual freedom, iii) Respect for democracy, justice and the rule of law, iv) Equality, non-discrimination, and solidarity and v) Citizens' rights. In the same document, basic ethical principles were also identified in AI systems: i) respect for human autonomy, ii) prevention of harm, iii) fairness iv) explicability.

²⁵European Parliament, "Civil Law Rules on Robotics", 2016 - https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html

²⁶European Commission, EU Declaration on Cooperation on Artificial Intelligence, 2018 -

<https://ec.europa.eu/jrc/communities/en/node/1286/document/eu-declaration-cooperation-artificial-intelligence>

²⁷High-Level Expert Group on Artificial Intelligence, ETHICS GUIDELINES FOR TRUSTWORTHY AI, 2019 - <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>

This led to a (non-exhaustive) list of social, individual, and sistemi requirements regarding the construction of reliable AI:

- a. **Human intervention and surveillance** (including fundamental rights, human intervention and human surveillance)
- b. **Technical robustness and security** (including attack resilience and safety, contingency planning and overall security, accuracy, reliability and reproducibility)
- c. **Confidentiality and data governance** (including respect for confidentiality, data quality and integrity, and access to data)
- d. **Transparency** (including traceability, explainability and communication)
- e. **Diversity, non-discrimination and equity** (including prevention of un fair bias, accessibility and universal design, and stakeholder participation)
- f. **Social and environmental well-being** (including sustainability and environmental compliance, social impact, society and democracy)
- g. **Accountability** (including verifiability, minimisation of adverse effects and their reporting, trade-offs and appeals)

These fundamental steps led to the publication in 2021 of the AI package, a set of new rules and actions proposed by the European Commission to turn Europe into the global hub for trusted AI. This package consists of:

- *A Communication on Fostering a European Approach to Artificial Intelligence*²⁸;
- The *Coordinated Plan with the Member States*²⁹: 2021 update;
- *A Proposal for a Regulation on AI*³⁰ that establishes harmonised rules for the EU (*Artificial Intelligence Act*).

The Artificial Intelligence Act

The Artificial Intelligence Act (AIA) was the European Commission's response to the need to create binding regulatory oversight of AI. In the European context, the AIA presents itself as the first binding step of the Union in AI. HLEG experts have osseste it as one of the most significant projects undertaken at the international level. Overall, it represents a starting point for ensuring that the development of AI in the EU is ethical, legal, socially equitable, and sustainable, with a vision of AI tha include economic, societal, and environmental aspects. In this sense, the project is very ambitious, so it will take time and effort to reach a final text that can achieve all the objectives. Indeed, itshould be borne in mind that the AIA Act is currently still a legislative proposal. The passage of this proposal into binding legislation will require confirmation by the European Parliament and the Council. This process could take a few years and is likely subject to negotiations, revisions, and compromises. As a result, the vision for AI that has been defined and outlined by the Commission could still be modified by MEPs, the influence of Member States through the EU Council, and, indirectly, the influence that the private sector could also exert.

Title I of the AIA defines the subject matter of the regulation and the scope of the new rules regarding the placing on the market, commissioning, and the use of AI systems and sets out the definitions used throughout the act. As pointed out in the research "*Achieving a 'Good AI Society'*":

²⁸European Commission, Communication on Fostering a European approach to Artificial Intelligence, 2021 - <https://digital-strategy.ec.europa.eu/news-redirect/709089>

²⁹European Commission, Coordinated Plan on Artificial Intelligence 2021 Review - <https://digital-strategy.ec.europa.eu/news-redirect/709091>

³⁰European Commission, 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 - <https://digital-strategy.ec.europa.eu/news-redirect/709090>

Comparing the Aims and Progress of the EU and the US" by Roberts et al. (2021)³¹, the definition of AI provided by the proposed European law is comprehensive and includes "statistical approaches." This will undoubtedly help to future-proof the EU definition. Still, it could also include systems that are not commonly considered part of AI, with a possible impact on the innovation sector. In particular, the draft law in Article 3(1) presents the following definition of AI:

- **Article 3(1)** - "Artificial intelligence system" (AI system) means software developed using one or more of the techniques and approaches listed in Annex I, which can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations or decisions that affect the environments with which they interact
- **Annex I** - Artificial intelligence techniques and approaches, referred to in Article 3(1): (a) Machine learning approaches, including supervised learning, unsupervised learning, and reinforcement learning, using a wide range of methods, including deep learning; (b) Logic-based and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inferential and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimisation methods.

In general, the framework outlined through the AIA approaches the AI regulation through a **risk-based system** which, taking into consideration the previous experiences of soft law and the work of the HLEG, will prohibit the AI systems considered to be at unacceptable risk. A multilevel system of regulatory requirements is presented in Titles II and III of the bill, depending on the inherent risk associated with the AI system or practices used.

- **Prohibited Artificial Intelligence Practices (Title II AIA):** The law aims to prohibit specific AI practices and systems deemed to generate an unacceptable risk. These include the use of AI by public authorities to obtain so-called 'social scoring' for assessing the behaviour of individuals, the use of real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes (subject to certain exceptions), and subliminal techniques aimed at significantly altering a person's behaviour in a harmful way.
- **High-risk AI systems (Title III AIA):** a specific category is established for high-risk AI systems, i.e., technologies that present a significant risk of causing harm and could negatively impact individuals' security or fundamental rights. For these reasons, their use is subject to a series of specific governance requirements that allow their use only in the presence of particular ex-ante security and compliance controls. The classification of an AI system as high-risk is based on its intended purpose, in line with existing EU product safety legislation. Consequently, classification as high-risk depends not only on the function performed by the AI system but also on the specific purpose and manner of use of that system. Chapter 1 of Title III lays down classification rules and identifies two main categories of high-risk AI systems: (i) AI systems intended for use as safety components of products subject to ex-ante conformity assessment by third parties, (ii) other stand-alone AI systems having implications primarily concerning the fundamental rights explicitly listed in Annex III. These include, for example, biometric identification systems both in real time and ex-post, credit scoring, AI systems related to essential public infrastructure, justice, social security, medical devices, and other regulated devices.

³¹Roberts, H., Cows, J., Hine, E. et al. Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US. *Sci Eng Ethics* 27, 68 (2021) - <https://doi.org/10.1007/s11948-021-00340-7>

- **Non-high-risk AI systems (Title III and Title IX AIA):** like the others, they will be subject to transparency requirements and encouraged to follow codes of conduct. Title IX establishes a framework for creating such codes of conduct to encourage providers of non-high-risk AI systems to apply the mandatory requirements for high-risk AI systems voluntarily. Suppliers will be able to develop and implement codes of conduct independently, and voluntary commitments relating to, for example, environmental sustainability, accessibility by persons with disabilities or diversity of development groups may also be included.

This risk-based framework and previous government initiatives seek to promote core European values by protecting fundamental rights, privacy, democracy, and human dignity while seeking to promote the improvement of economic performances and minimise social risks. The ethical principles already defined by the HLEG regarding accountability and transparency requirements have also been included in the AIA. This action has been judged positively by the experts for the ethical development of the AI. Specifically, the transparency requirements outlined in Title IV will apply to systems that: i) interact with humans, ii) are used to detect emotions or establish an association with (social) categories based on biometric data, iii) generate or manipulate content ("deep fakes"). In general, AI systems will need to be designed and developed to ensure that their operation is sufficiently transparent to allow users to interpret the system's output and use it appropriately. People will have to be informed when they will have to interact with an AI system and when their emotions or characteristics will be recognised through automated means.

To prevent the application of ethical principles in AI from affecting innovation processes, measures to create a legal framework conducive to innovation adapted to future needs and resilient to disruption have also been included in Title V of the bill. Competent national authorities are encouraged to create space for regulatory experimentation, defining a basic governance, control, and accountability framework and reducing some regulatory burdens for SMEs and start-ups.

Instead, as far as governance is concerned, the AIA provides in Titles VI, VII, and VIII, the creation of the *"European Artificial Intelligence Board"* to offer consultancy and assistance to the Commission on the subjects treated by the bill proposal. In particular, the Board will contribute to practical cooperation between national supervisory authorities and the Commission and ensure consistent application of the rules. In addition, the structure of the Board will be modelled on the powers and responsibilities of the *European Data Protection Board (EDPB)* under the General Data Protection Regulation (GDPR)³². Member States will also be required to designate competent national authorities and national supervisory authorities to provide guidance and advice on implementing the rules.

Finally, Titles X, XI, and XII contain the 'Final Provisions'. Title X emphasises the obligation for all parties to respect the confidentiality of information and data and sets out the rules for the exchange of information obtained during the implementation of the Regulation. Title XI lays down the rules for delegation and implementing powers, empowering the Commission to adopt implementing acts to ensure the uniform application of the rules. Title XII instead contains the obligation for the Commission to periodically update the requirements for the assessment of an AI system, also providing for a differentiated transitional period for the initial date of applicability of the law to facilitate the correct implementation of the rules for all parties.

Comments on the strategy of the European Union

³²European Parliament and Council, General Data Protection Regulation, 2016 - <https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32016R0679>

Overall, the European Union's long-term strategy for an ethical vision of AI, including the mechanisms for affirming it, was considered consistent by the experts. The 'risk-based' system to be introduced with the Artificial Intelligence Act aims to protect the individual rights of European citizens, thereby promoting collective welfare and civil society³³. The three fundamental pillars of the AIA, already outlined by the Commission in the 2018 *Communication on the European Approach to AI*, are:

- 1) **Improving economic performance**
- 2) **Minimising social disruption**
- 3) **Develop appropriate ethical and legal frameworks**

The 'risk-based' approach, combining hard and soft law elements, aims to ensure that harm to people is minimised while allowing for the commercial and social benefits of these technologies³⁴. Nevertheless, the main criticism that has been levelled at the European AI vision is that it focuses too much on protecting individual rights and does not sufficiently stimulate innovation, which could hamper economic growth and competitiveness, according to some experts³⁵.

About the first two general priorities of the law, i.e., increasing the industry's capacity and preparing for social disruption, several initiatives have been introduced, with a joint effort by the Commission and the Member States to invest in the opportunities presented by AI. For example, the European Investment Bank (EIB) and the European Investment Fund (EIF) have pledged EUR 150 million to support AI companies across Europe³⁶. Many policies have also been outlined that seek to mitigate the social disruption that AI can cause, including the *'Digital Education Action Plan 2021-2027'* to ensure that citizens have strong digital skills and the introduction of a code of conduct on misinformation platforms to follow. As for the third priority, developing legal and ethical frameworks for AI governance, the first important step was taken with the ethical recommendations of the HLEG, a non-binding guide whose principles have been largely taken up in the Artificial Intelligence Act.

From a regulatory perspective, the proposed AI Act provides a solid basis for standardising protections across Europe, banning certain use cases, and giving standard criteria for defining and regulating high-and limited-risk AI. Even if enforcement is relatively standardised across the EU, there is a risk that the proposals could '**over-regulate**' or '**under-regulate**' AI.

Over-regulation. The proposed AI bill states that "*certain AI systems intended to distort human behaviour, where by physical or psychological harm is likely to occur, should be banned in terms of banned systems.*" This clause is inclusive and could potentially include several use cases, such as recommender systems that are intended to push an individual towards a specific type of content while limiting their exposure to others. Recommender systems can undoubtedly be harmful, but prohibiting them would be a disproportionate response, given the benefits they bring to daily life when using digital services, such as buying products, listening to music, or watching movies online.

³³Roberts, H., Cows, J., Hine, E. et al. Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US. *Sci Eng Ethics* 27, 68 (2021) - <https://doi.org/10.1007/s11948-021-00340-7>

³⁴The IEEE Initiative on Ethics of Autonomous and Intelligent Systems. (2017). *Ethically Aligned Design*, v2. Retrieved September 18, 2018 from <https://ethicsinaction.ieee.org>

³⁵Brattberg, E., Csernaton, R., & Rugova, V. (2020). Europe and AI: Leading, lagging behind, or carving its own way? Carnegie Endowment for International Peace. <https://carnegieendowment.org/2020/07/09/europe-and-ai-leading-lagging-behind-or-carving-its-own-way-pub-82236>

³⁶European Commission. (2020b). New EU financing instrument of up to €150 million to support European Artificial Intelligence companies. <https://digital-strategy.ec.europa.eu/en/news/new-eu-financing-instrument-eu-150-million-support-european-artificial-intelligence-companies>

Moreover, without further clarification, over-regulation could stifle innovation through increased costs and red tape, leading the EU to lose competitiveness against other countries.

Under-regulation. At the same time, questions may be raised about EU under-regulation. Some HLEG members have criticised the vague and non-binding nature of the guidelines due to possible influence on the design of the proposed legislation by leading industry groups, which would have been more robust than the action of the HLEG experts. The AI Act may alleviate the concerns of non-binding enforcement, but weaknesses in the measures remain. The ban on 'real-time' remote biometric surveillance has key exclusions, including for cases of missing children or terrorist threats. The European Data Protection Supervisor, *Wojciech Wiewiórowski*, said he was disappointed that the law did not include a complete moratorium on remote biometric surveillance³⁷. Similarly, the text surrounding so-called 'disparate impact' assessments was vague and non-committal, with few formal requirements for bias checks. As a result, adequate protection from high-risk systems will largely depend on interpretations by standardisation bodies and effective internal compliance by companies, which could lead to ineffective or unethical results in practice.

ANNEX I

ITALY

Proposals and initiatives

Reference to:

- a) key government documents and plans
- b) trade union recommendations

In Italy, the last few years have seen a vast production of documents that have tried to technically define what Artificial Intelligence-AI is in order to provide useful elements for its legal framework and adequate regulation. A significant element to keep in mind concerns the fact that often the documents drawn up by public institutions are the result of an active collaboration with private subjects. This annex presents the most significant public acts of the Italian system and the synthesis of significant proposals and recommendations drawn up by the trade union in the world of work. .

The AI White Paper -2018

In 2012, the Monti government established the **Agency for Digital Italy (AgID)**³⁸, a public agency with the task of pursuing the highest level of technological innovation in the organisation and development of public administration, improve services to citizens and businesses, in compliance with legality, impartiality, transparency, and efficiency criteria. On 21 March 2018, AgID officially presented the AI White Paper³⁹, a document edited by the AI Task Force of the Agency and produced through a work of consultation, synthesis and in-depth analysis that involved about a hundred public and private entities that in various capacities deal with AI in Italy.

This paper is the first Italian document addressed to public administrations (schools, health facilities, municipalities, courts, ministries). It contains recommendations and indications on how to

³⁷European Data Protection Supervisor. (2021). Artificial Intelligence act: A welcomed initiative, but ban on remote biometric identification in public space is necessary. https://edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative_en

³⁸ Agenzia per l'Italia Digitale - <https://www.agid.gov.it/>

³⁹ Agenzia per l'Italia Digitale, Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino, a cura della Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale, marzo 2018 - <https://libro-bianco-ia.readthedocs.io/it/latest/>

make the most of the opportunities offered by Artificial Intelligence, limiting its critical and problematic aspects and developing citizen-friendly public services. The paper addresses and analyses the various challenges for the future, the ethical, legal, technological, and cultural dynamics related to the relationship of new technologies with human beings, and those arising from implementing AI in public administration, such as eliminating inequalities, measuring impact, and accompanying transformation, following a multidisciplinary and systemic approach for all of them. Moreover, the paper contains a decalogue of recommendations for the public sector, with a series of indications that aim to respond to the challenges addressed in the document.

The “Italy 2025 Plan” - 2019

In December 2019, the *Italy 2025 Plan (Piano Italia 2025)*⁴⁰ was presented by the Minister for Technological Innovation and Digital Transition as part of the initiatives for the digitisation of the country. The Plan represents an overall strategy that mainly points to three challenges: digital society, the goal of innovation, and sustainable and inclusive development. Specifically, concerning the third challenge, it is reported that “the state governs the use of technologies, promoting their diffusion, maximising collective benefits and minimising negative impacts. The search for ethicality, responsibility, and non-discrimination of technological solutions, especially in Artificial Intelligence, becomes a guiding star of government action”.

The 20 main innovation actions to address the challenges outlined in the Plan are the following:

1. A01 - A governance for innovation and digital
2. A02 - Digital identity (reloaded)
3. A03 - A digital domicile for all
4. A04 - IO, the public services app
5. A05 - Digital Restructuring
6. A06 - Open innovation in public administration
7. A07 - Simplified procurement for innovation
8. A08 - Artificial intelligence at the service of the State
9. A09 - Data for the cities of the future
10. A10 - Villages of the future
11. A11 - Innovation as a common good
12. A12 - Right to Innovate
13. A13 - Made.IT, from idea to innovative enterprise
14. A14 - Cross-Tech hub Italy
15. A15 - MoonTransfer Fund & Missione Training
16. A16 - Shared, secure, reliable and green digital infrastructures
17. A17 - AI ethical LAB-EL
18. A18 - The Saturday of the future
19. A19 - An older person, a tablet and a smile for digital inclusion
20. A20 - Digital Republic: a training hub on the future

In particular, point A17 envisaged the creation of an "Alliance for Sustainable Artificial Intelligence", “a committee to which public and private actors will be invited to join. The committee's first task will be to draw up an ethical-legal statute for artificial intelligence based on the results of the various groups of national and European experts who have already addressed the issue. This statute will establish a set of minimum guiding principles as well as rules for the qualification of artificial intelligence solutions intended for the public and private sectors, a sort of

⁴⁰ Agenzia per l'Italia Digitale, Strategia per l'innovazione tecnologica e la digitalizzazione del Paese 2025, dicembre 2019 - <https://docs.italia.it/italia/mid/piano-nazionale-innovazione-2025-docs/it/stabile/index.html>

certification of the ethical-legal sustainability of the solution that could then be translated into a certificate of a passed ethical impact assessment on society”.

The Memorandum of Understanding on Artificial Intelligence Ethics in Public Administration - 2020

On 18 February 2020, Minister for Technological Innovation and Digitalisation *Paola Pisano* and Leonardo Foundation President *Luciano Violante* signed a memorandum of understanding⁴¹. The Memorandum defines the ethical and legal framework within which to develop and apply artificial intelligence, in particular, to meet the needs of public administration. The primary objective of the document is based on the desire to introduce artificial intelligence applications in the management of administrative procedures with the 'commitment to put man at the centre, promoting "socially, culturally and democratically sustainable artificial intelligence". The collaboration born with the protocol specifically envisages:

- the definition of an assessment methodology that can guarantee, during the design, development and implementation phases, the sustainable use of AI in public services while respecting constitutional values of Italian republic;
- the drafting of a proposal for a "compliance code" for the implementation of AI in the public or private sector, also to define a certification system for ethical and legal sustainability;
- the definition of a training plan for school teaching staff on the basic concepts and methods of AI, starting from the analysis of benefits and risks to the rules of conduct for "beneficial" AI;
- the definition of at least two projects intended for possible experimentation, dedicated to applying AI in administrative and judicial proceedings. These projects will also be identified in the context of the inter-ministerial steering committee for the country's innovation, which was set up by the Minister for Technological Innovation and Digital Transition.

National Research Programme 2021-2027

On 15 December 2020, the Inter-ministerial Committee for Economic Planning - CIPE approved the National Research Programme 2021-2027⁴², the result of a debate initiated by the Ministry of Universities and Research with the scientific community, state and regional administrations, and extended to public and private stakeholders and civil society. The document takes the form of a multi-year framework programming tool designed to contribute to the achievement of the UN *Sustainable Development Goals (SDGs)*, the priorities of the European Commission, the 2021-2027 Cohesion Policy Goals, as well as the *Next Generation EU* initiative. The application areas of the Programme are the following:

- Health
- Humanistic culture and inclusion
- Security for social systems
- Digital, industry and aerospace
- Climate, energy, sustainable mobility
- Food, natural resources, agriculture, environment.

For each research area, the relevant thematic domains are formulated. In the area of AI, the Programme envisages several major domains of action and, for the first time, a specific 'Artificial Intelligence' domain in close coordination with other sectors, such as digital transformation, Big

⁴¹Ministro per l'innovazione tecnologica e la transizione digitale, Un protocollo d'intesa sull'etica dell'Intelligenza Artificiale nella Pubblica Amministrazione, febbraio 2020 - <https://innovazione.gov.it/notizie/articoli/un-protocollo-d-intesa-sull-etica-dell-intelligenza-artificiale-nella-pubblica-amministrazione/>

⁴²Ministero dell'Università e della Ricerca, Programma nazionale per la ricerca 2021-2027, dicembre 2020 - <https://www.mur.gov.it/sites/default/files/2021-01/Pnr2021-27.pdf>

Data, robotics and cyber security. In particular, there are five articulations into which the NRP 2021-2027 is divided with regard to Artificial Intelligence⁴³:

1. **AI for AI:** "Italy must invest in research and development in Artificial Intelligence, as a scientific discipline: it must invest in foundational research in AI to maintain its current international scientific leadership; it must invest, also using a major National Project in AI, to impact with important contributions in all the articulations of the PNR, and to guarantee growth opportunities for Italian industry and a future of work for the new generations".
2. **AI for the person and health:** "modern AI is increasingly defining itself as a tool aimed at the person, both in terms of the individual's enhanced cognitive and predictive capabilities to support individual decisions and behaviour, and in terms of Human-AI interaction and, finally, to improve the person's well-being, health and life".
3. **AI for society:** "AI, through the processing of large amounts of data, shows enormous potential in responding to society's challenges, in areas ranging from its well-being to its planning, economic and scientific evolution".
4. **AI for the environment and critical infrastructures:** "environmental problems are, by their very nature, complex and multifaceted due to a large number of variables involved, their mutual interactions and the dominant presence of uncertainty also caused by interaction with humans. Creating a model capable of computationally describing such complexity (or parts of it) is a challenge that AI can take up and overcome".
5. **AI for industrial production:** "AI is the lynchpin of the new industrial devolution complementing the paradigms of Industry 4.0, which goes in the direction of adopting, customising and co-creating intelligent systems for process optimisation and the design of new generations of AI-native products'. The research will have to cover quality control and diagnostics, the search for new molecules, the design of new materials, the creation of Digital Twins for the simulation of complex industrial scenarios, and the development of edge-AI systems that can be integrated into production and products, and so on."

Strategic Programme for Artificial Intelligence (AI) 2022-2024

On July 2021, the Ministry of Universities and Research, the Ministry of Economic Development and the Minister for Technological Innovation and Digital Transition set up a working group of experts with the task of supporting the ministries in updating the national strategy on Artificial Intelligence, in particular to make it consistent with the **National Recovery and Resilience Plan - PNRR** and recent developments at EU level. On 24 November 2021, Italy then adopted the Strategic Programme for Artificial Intelligence (AI) 2022-2024⁴⁴. This Strategy defines a coherent and holistic framework of initiatives to support the development of a national AI ecosystem. Its design is inspired by five guiding principles:

1. Italy's AI is a European AI. In line with the EU Coordinated Plan on Artificial Intelligence, the Italian Strategic Programme stems from the awareness that only through common and coordinated actions Europe will be able to compete globally and work towards strategic autonomy. Therefore, this programme reflects the four sets of proposals put forward by the 2021 EU Coordinated Plan on AI. First, it sets enabling conditions for AI's development and uptake by focusing on cooperation, data and computing infrastructure. Second, it

⁴³Innovation Post, "Programma Nazionale di Ricerca, aperta la consultazione pubblica: Industria 4.0 e robotica protagoniste", di Francesco Bruno, 21 agosto 2021 - <https://www.innovationpost.it/2020/08/21/pnr-2021-2027-consutazione-aperta/>

⁴⁴Governo Italiano, a cura del Ministero dell'Università e della Ricerca, del Ministero dello Sviluppo Economico e del Ministro per l'Innovazione tecnologica e la Transizione Digitale e del gruppo di lavoro sulla Strategia Nazionale per l'Intelligenza Artificiale, Programma Strategico Intelligenza Artificiale 2022-2024, luglio 2021 - <https://assets.innovazione.gov.it/1637937177-programma-strategico-iaweb-2.pdf>

leverages on existing Italian High Performance Computing-HPC and data-management infrastructure. Third, it aims at nurturing talents and adheres to the joint effort for improving and adopting the harmonised set of rules for AI proposed by the EU AI Act. Fourth, it identifies priority sectors where to build strategic leadership.

2. Italy will be a global research and innovation hub of AI. To guarantee future economic growth and strategic autonomy, it is essential for Italy to boost its AI research and development ecosystem to generate cutting-edge AI innovations. Accordingly, this strategic programme will invest in frontier research and applications to develop AI methodologies and techniques of tomorrow.
3. Italy's AI will be human-centred, trustworthy and sustainable. Technologies must not promote economic growth *per se*, but inclusive and sustainable growth, in line with the principles contained in Article 3 of the Italian Constitution and the United Nations Sustainable Development Goals. This means that AI development must be centred around economic and social inclusion, human rights as well as environmental sustainability. AI must be designed and implemented in a responsible and transparent manner, based on trust and robustness so that it can be safely adopted in every sector and be capable of responding to societal challenges. To this aim, Italy adheres to the "Ethics Guidelines for trustworthy AI- Guidance and implementation program" defined by the High Level Expert Group on AI, inside EU.
4. Italian companies will become leaders of AI-based research, development and innovation. The digital transformation of the Italian entrepreneurial ecosystem is a must, if Italy wants to keep up with the most developed and innovative nations. To that end, this programme fosters the development, implementation and adoption of AI solutions. Public-private partnerships will be instrumental in finding appropriate sinergie between research bodies and enterprises with the aim of increasing Italy's technology transfer capabilities and thus competitiveness.
5. Italy's public administrations will govern with AI and will govern AI. The use and impact of AI in the public sector revolves around the dual dimensions of governance "*with and of*" AI. On the one hand, Italy's Government will improve its internal processes and policies thanks to a responsible use of data and AI technology. On the other hand, the Government is committed to governing AI and mitigating its potential risks, especially to safeguard human rights and ensure an ethical deployment of AI.

Moreover, this strategic programme lays out six objectives with a view to boosting Italy's strengths while mitigating its weaknesses. These objectives are:

- **Objective 1.** Advance frontier research in AI, both fundamental and applied, with a view to generating impact on important priority sectors of Industry, Public sector, Society, and Environment. A multidisciplinary approach must be incentivized, where research coexists with industrial and social innovation generating true innovation ecosystems.
- **Objective 2.** Reduce AI research fragmentation by helping the AI ecosystems achieve critical mass and by fostering network collaborations, making scientific excellence coexist with social cohesion and territorial inclusiveness.
- **Objective 3.** Develop and adopt human-centred and trustworthy AI in the public and private sector, promoting the societal acceptability of AI solutions, their compliance to regulation of AI, while supporting the development and design of responsible AI technology and systems.
- **Objective 4.** Increase AI-based innovation and the development of AI technology by fostering industrial investments and partnerships that drive excellent science out to the market, hence facilitating the uptake of AI in Small and Medium Enterprises - SMEs.
- **Objective 5.** Develop AI-driven policies and services in the public sector by boosting public sector innovation, the adoption of AI solutions and the cooperation between research centres, industries, and public bodies.

- **Objective 6.** Create, retain and attract AI talent in Italy by promoting all levels of education in AI, creating a new generation of holistic AI researchers and innovators, making Italy an attractive AI destination for qualified human capital from abroad, with a particular attention to diversity and gender balance.

The “Digital Republic”: the National Strategy for Digital Skills and the related Operational e-Skills Plan – 2020, 2021

Within the framework of the Italy 2025 Plan, the last of the 20 points outlined (A20) envisaged the creation of the project “Repubblica Digitale (Digital Republic): a training hub for the future”.⁴⁵, an initiative with the aim of fighting digital divides and educating on the technologies of the future. In particular, “Repubblica Digitale” was born “from the idea that accompanying the country's digital transformation with informative, educational and training actions is a civic duty of the State, businesses and individual citizens. All the components of society, starting with the media, the school and the family, must make a decisive contribution to overcoming the gaps that characterise our country”.⁴⁶ The aim of the project is therefore to combat the digital divide of a cultural nature present in the Italian population, to support maximum digital inclusion and to foster education on the technologies of the future. To do this, the initiative aims to create an alliance between public and private entities/organisations together with citizens, based on the common recognition of principles and needs and on the common will to face the digital challenge with an ethical spirit. Already today, “Repubblica Digitale” has seen the participation of large companies, consumer associations, public administrations and non-profit organisations, united in the *National Coalition for Digital Competence* (composed of the public and private entities that adhere to the Manifesto of the project). The Manifesto⁴⁷, in particular, lists three fundamental commitments:

1. **Digital education:** IT culture and digital skills are essential requirements of citizenship. The public and private sectors must invest resources in their development as determining factors for the country's growth, competitiveness and well-being, combating all forms of digital illiteracy also through schools, universities and the media.
2. **Digital citizenship:** digital technology can foster the development of a new form of citizenship based on quality information, participation in deliberations, civic interaction and a more effective relationship between citizens and public administration. Digital technology designed with a focus on citizens' rights can become the common language in the dialogue between citizens, public administrations and businesses and contribute to overcoming inequalities. Public and private make their services available digitally in an accessible and citizen-friendly manner without creating new technological barriers and by breaking down existing ones.
3. **Ethical, human and non-discriminatory digital:** digital can become a space for equality and development of communities and individuals. Public and private contribute to the elimination of all social, economic, geographical, technological and cultural barriers that may prevent de facto equality between citizens in the use of public and private digital services and in accessing the opportunities offered by digital. Digital must be ethical and anthropocentric. Public and private must design, develop and deliver their services in accordance with the fundamental ethical values recognised in our country and the rest of the world, placing the person and the community at the centre.

⁴⁵Ministro per l'innovazione tecnologica e la transizione digitale, “Repubblica Digitale. Una risposta organica e di sistema per lo sviluppo delle competenze digitali quali asset strategico per il Paese”, 2021 - <https://innovazione.gov.it/progetti/repubblica-digitale/>

⁴⁶<https://docs.italia.it/italia/mid/piano-nazionale-innovazione-2025-docs/it/stabile/piano-d-azione/A20-repubblica-digitale.html>

⁴⁷Ministro per l'innovazione tecnologica e la transizione digitale, “Il manifesto per la Repubblica Digitale” - Il manifesto per la Repubblica Digitale

Furthermore, in order to make digital a possible reality for social and economic growth, systemic actions have been defined, included in the *Operational Plan of the e-Skills Strategy*⁴⁸. The Plan, published in December 2020, contains the measures deemed necessary to bring down digital illiteracy and develop a pathway that can invest all sectors of society. The main objectives that the Plan aims to achieve by 2025 are as follows:

- raise the share of the population with at least basic digital skills to 70%, an increase of more than 13 million citizens compared to 2019, and close the gender gap to zero;
- double the population with advanced digital skills;
- triple the number of ICT graduates and quadruple the number of female graduates; double the share of companies using big data;
- increase the share of SMEs using ICT specialists by 50%;
- increase the proportion of the population using public digital services by five times, reaching 64%. To bring Internet use up to the levels of the most advanced European countries, even in the younger segments of the population (to 84% in the 65-74 age group).

Essential for the achievement of the objectives is the activity of the *National Coalition for Digital Competencies*, which has the task of promoting concrete actions, capable of producing measurable and quantifiable results. It should also be noted that, in order to strengthen the projects already underway, the Minister for Technological Innovation and Digitalisation signed the decree adopting the National Digital Competence Strategy in August 2020.⁴⁹, the first example of an e-skills strategy in Italy. This Strategy envisages four axes of intervention:

1. the development of the digital skills needed within the higher education and training cycle, with the coordination of the Ministry of Education and the Ministry of University and Research;
2. the enhancement and development of the digital skills of the workforce, both in the private and public sectors, including e-leadership skills, with the coordination of the Ministry of Economic Development and the Ministry of Public Administration
3. the development of specialised ICT skills for new markets and new jobs, largely related to emerging technologies and the possession of key skills for the jobs of the future with the coordination of the Ministry of Universities and Research and the Ministry of Economic Development;
4. the enhancement of digital skills necessary for exercising citizenship rights (including the full use of online services, which is particularly necessary at this time) and informed participation in democratic dialogue under the coordination of the Minister for Technological Innovation and Digitisation.

Actions for digital ethics from the civil society - 2022

It is clear that digital skills are to be thought of as a necessary condition for those who work, as well as for being citizens, for example: by exploring the relationship between the level of awareness and consequences (for company development, for career development, for the defence of rights, etc.); with actions oriented towards the national strategy (e.g. transversal training with the inter-professional funds); by raising the issue of choosing the direction of change and thus spreading the culture of e-leadership even in the branches of trade unions. In particular, the 20 January 2022 an event organised by the Italian trade union CISL, *"The Platform Society. Social networks, fake news and conspiracy: origins, dissemination and social repercussions of a phenomenon that also entails risks for trade unions"*, focusing on the damage to the social fabric that fake news and conspiracy (consequences, in fact, of a lack of digital awareness) can cause. Fake news, in particular:

⁴⁸Ministro per l'innovazione tecnologica e la transizione digitale, Piano Operativo della Strategia per le competenze digitali - <https://repubblicadigitale.innovazione.gov.it/it/le-azioni/#documenti>

⁴⁹Ministro per l'innovazione tecnologica e la transizione digitale, Strategia Nazionale per le Competenze Digitali - <https://repubblicadigitale.innovazione.gov.it/italia-ha-sua-strategia-nazionale-competenze-digitali/>

- Identify an enemy and a secret plot that threatens citizens' lives or beliefs, and trigger a defence mechanism that can fuel discrimination, justify hate crimes, and provide pretexts for violent extremist groups.
- Spread mistrust in institutions, which can lead to indifference or political radicalisation, and distrust, if not outright hostility, towards parties and trade unions. The risk of a violent drift, as the case of the storming of Capitol Hill by hundreds of supporters of former US President Donald Trump shows, is obvious.
- Spread mistrust of scientific and medical information, with the risk of severe consequences for public health.
- Harm those who fight against the real mysteries and lies of power. The risk is that even those who investigate the true dark sides of history end up being dismissed as a conspiracy theorist and his work discredited.
- Create the preconditions for an authoritarian and undemocratic drift. These isolated individuals perceive that they are under siege, so the most obvious solution seems to be to rely on someone who knows what's what. These strong personalities claim to be able to find the culprit of the problems and have simple turnkey solutions to resolve the situation.

In fact, digitalisation is an important aspect when talking about work and job search, which is why the union's commitment in this area is of great importance. Specifically, the CISL has outlined the following proposals, which are the result of intense discussion within the Coordination working group:

The regulatory framework and shared governance of the Algorithm: the union must set itself the objective of participating and contributing at the various existing levels of interlocution - global, European, national, sectoral and corporate - to the construction of a regulatory framework that protects the fundamental rights and freedoms of individuals and their personal data. In general, the union highlights the need to:

- a) increase the taxing power of states in which multinational companies operate and combat monopolies in the high-tech sector;
- b) support all initiatives aimed at designing, testing and building new media platforms and environments to foster info-diversity in the digital system;
- c) introduce and apply new economic and tax mechanisms to make the business model adopted by Big Tech more fair and sustainable. If, as is evident, data are the real wealth of these companies - thus the union's motivations - and if data, as scholars on the matter confirm, are the new oil of the 21st century, then it does not seem improper to propose also a payment of royalties on the data acquired by these companies to be allocated to a common fund aimed at countering the adverse effects of the digital transition in terms of increasing inequalities, new social marginalities, increase unemployment and precariousness, disinformation and educational poverty;
- d) d). promote research and training activities and information campaigns (which can be financed with the resources of the fund describe above) to increase platform users' awareness of the use and value of their data. The issue of misinformation and combating the phenomenon of online social hatred should also see the union at the forefront of a coordinated awareness campaign starting in the workplace to encourage online living conditions that are more respectful of human rights, dignity, life and integrity of people.

On a regulatory level, the trade union CISL proposes introducing a **Digital Code of Ethics** in Italy, operating consistently with the declaration approved by UNESCO-IFAP in 2021, the *Ugra Memorandum* (III International Conference, Khanty-Mansiysk (RF), 17–18 June 2021). The primary need is to adopt a Universal Digital Code of Ethics, where the reference values and

principles are defined to steer digitalisation processes toward honest and fair political, economic, social and cultural progress of humanity.

Furthermore, it believes that to combat disinformation and fake news, **specific regulations should be introduced on the responsibility of social platforms** as vehicles of false information or information detrimental to dignity, health, civil coexistence, or the integrity of the human person. To this end, it also proposes the establishment of super partes control and moderation bodies to protect both freedoms of expression and information and respect for users' fundamental rights.

The trade union document notes that at both European and national legislative levels, steps forward have been taken in this direction with the *General Data Protection Regulation - GDPR Regulation No. 2016/679* - and that more is being attempted with the latest proposals on digital services and markets as well as on artificial intelligence by the European Commission. However, there is still much to be done, the union comments, especially on rights and collective action. One example: the provisions of the GDPR were introduced to protect the fundamental rights and freedoms of individuals and, to this end, prohibit certain practices, or allow them only with certain limitations [Articles 22 and 35 GDPR]. But the rights that are to be protected are within the sphere of the individual, so much so that Article 88 of the GDPR - the so-called 'passerelle rule' - suggests the possibility for member states to provide by law or through collective agreements for further, **more specific rules for employees in the context of labour relations**; a new field of legislative and contractual regulation that is still largely to be explored.

Regarding governance, the trade union, like the leading development players, should work towards constructing joint prevention and monitoring mechanisms at corporate, sectoral, national, European and global levels. In particular, the union emphasises the need for monitoring the impact of technologies concerning organised structures and, consequently, on the conditions of the rights of the people involved. National regulation is currently a complex issue because company situations are very dynamic. The trade union should define a general policy and then apply it in specific cases, implementing an improved new model for new technologies.

The digital skills necessary for workers and trade unionists: digital skills and digital awareness thought as necessary conditions for workers and citizens. In this regard, the CISL document proposes a targeted study on the digital skills of workers concerning a plurality of dimensions (size and technological intensity of companies, territory, level of responsibility/designation) and related to the effects produced (for the development of companies, for career development, for the defence of rights). The objective is to raise awareness of the transformation taking place and to implement more targeted interventions in favour of workers and within the union itself. Training aspects are preparatory to any action in this area. It is necessary to provide extensive training courses for trade unionists in the first place, but then, through the use of interprofessional funds, also for workers within companies on the processing and use for management purposes of personal data, on the management of business processes based on the algorithm, on automated decision-making mechanisms and profiling and their possible impact on workers' lives.

Ad hoc provisions in the collective agreement: for the trade union, it is essential to provide some ad hoc provisions in collective agreements for the appropriate supervision of digital processes and technologies introduced/to be introduced in the company. For example, within the so-called 'bilaterality' framework, it would be appropriate to provide for a *bilateral commission on AI and digitalisation* at the level of the collective agreement with the tasks of information and consultation, monitoring, and impact assessment for the introduction of new digital technologies. Another suggestion is the provision to introduce in companies, at the production unit level, the figure of an ad hoc trade union manager on AI and digitalisation concerning which to imagine a two-way flow of information both with workers and with the company data protection officer.

On these issues, in March 2022, the union participated in the work of the Partnership Table envisaged by the government's Reconstruction and Resilience Plan - PNRR and in the meeting with the Minister for Technological Innovation and Digital Transition⁵⁰.

Finally, the contribution of the national business association Confindustria, which created **Confindustria Digitale**⁵¹ in 2011, is also worth mentioning. This Federation of Industrial Representatives was created with the aim of promoting the development of the digital economy, to the benefit of the country's competition and innovation and, through its *Economic Policy Review*, publishes expert contributions on the dynamics and challenges of digital transformation⁵².

ANNEX II

GERMANY

Proposals and initiatives

Reference to:

- a) the "Strategy for automated and connected driving"
- b) the government's strategy for AI
- c) the approach of the President of Germany F.W. Steinmeier on the relationship between Ethics and Innovation

The "Strategy for Automated and Connected Driving" - 2015

In 2015, the German government officially adopted the 'Strategy for Automated and Connected Driving - Remain a leading provider, become a leading market, introduce regular operations'.⁵³, based on the work of the Federal Ministry of Digital and Transport, which looks at the context of automated driving as a source of numerous priorities for the country's development. A better connection of the different modes of transport through digital solutions is something that Germany considers central in the area of digital mobility. The Strategy provides a regulatory framework for automated driving, and, within it, special attention is paid to data protection and data security. In recent years, this strategic plan has been implemented through targeted measures in the action areas of infrastructure, legislation, innovation financing, connectivity, IT security, data protection and social dialogue. The main results of the federal government's implementation of the strategy were summarised in a report⁵⁴ at the end of the 18th parliamentary legislature and were:

- the adaptation of the national legal framework, in particular the Road Traffic Act;
- the adoption of an action plan for the creation of ethical rules for self-driving computers;
- the establishment and coordination of test beds for automated and connected driving in the real world (Digital Test Beds);
- support for research and development of automated and connected driving solutions, from basic to application research;

⁵⁰CISL, "Tavolo partenariato PNRR – incontro con il Ministro per l'innovazione tecnologica e la transizione digitale", 21 marzo 2022 - <https://www.cisl.it/notizie/notizie-lavoro/tavolo-partenariato-pnrr-incontro-con-il-ministro-per-linnovazione-tecnologica-e-la-transizione-digitale/>

⁵¹Confindustria Digitale - <http://www.confindustriadigitale.it/federazione/profilo.kl>

⁵²Confindustria, Rivista di Politica Economica - https://www.confindustria.it/home/centro-studi/rivista-di-politica-economica/dettaglio?doc=RPE_digitale_2020_1

⁵³Federal Government, Strategy for Automated and Connected Driving, 2015 -

<https://www.bmvi.de/SharedDocs/EN/publications/strategy-for-automated-and-connected-driving.html>

⁵⁴ <https://www.bmvi.de/EN/Topics/Digital-Matters/Automated-Connected-Driving/automated-and-connected-driving.html#:~:text=Automated%20and%20connected%20driving%20is,for%20automated%20and%20connected%20driving.>

- active creation of rules and standards in European and international bodies.

The Act on Automated Driving - 2017

With the Act on Automated Driving⁵⁵ (Eighth Act amending the Road Traffic Act), which came into force on 21 June 2017, Germany became the first country in the world to regulate the rights and obligations of drivers using automated driving functions. This legislative change provided the necessary legal certainty for consumers as well as industry. If the regulatory requirements are met, drivers, taking into account the corresponding legal provisions, can divert their attention from the traffic environment and the operation of their vehicles while using automated driving systems in accordance with the law. Among other things, the law defines the high technical requirements that automated systems must meet to allow drivers to divert their attention. The relevant vehicle manufacturer must confirm that a system meets these technical requirements. If drivers comply with the new requirements and use the system in an proper, orderly manner, they can rely on the automated driving function working.

The Act on Autonomous Driving

Following the adoption of the law on automated driving, Germany now intends to enable driverless autonomous driving in certain areas of operation⁵⁶. The required legal framework is currently being prepared and will serve to allow motorvehicles with autonomous driving functions to use public roads in Germany until there are general harmonised provisions at international level.

The Independent Ethics Commission on Automated and Connected Driving and the “Action Plan for creating ethical rules for self-driving computers” - 2017

The *Independent Ethics Commission on Automated and Connected Driving*, established by the Federal Ministry of Digital and Transport in 2016, is a body comprising several distinguished experts from academia, society, the automotive industry and the digital technology sector to investigate ethical and socially relevant issues regarding automated and connected vehicle traffic. In 2017, the Commission produced a Report⁵⁷ that served as a basis for the government to adopt the ‘*Action Plan for creating ethical rules for self-driving computers*’.⁵⁸ The measures contained in this plan are currently being implemented. For example, on the initiative of Germany and in the context of the regularly held high-level structural dialogue on automated and connected driving, the EU Member States agreed to continue the discussion on cross-cutting ethical issues at the EU level to develop a European harmonised framework. The Commission report included, specifically, 20 rules:

1. The primary purpose of partly and fully automated transport systems is to improve safety for all road users. Another purpose is to increase mobility opportunities and to make further benefits possible. Technological development obeys the principle of personal autonomy, which means that individuals enjoy freedom of action for which they themselves are responsible.

⁵⁵Federal Government, Automated Driving Act, 2017 - <https://www.loc.gov/item/global-legal-monitor/2017-02-09/germany-government-proposes-automated-driving-act/#:~:text=Article%20Germany%3A%20Government%20Proposes%20Automated,be%20used%20on%20public%20roads.>

⁵⁶<https://www.bmvi.de/EN/Topics/Digital-Matters/Automated-Connected-Driving/automated-and-connected-driving.html#:~:text=Automated%20and%20connected%20driving%20is,for%20automated%20and%20connected%20driving.>

⁵⁷Federal Government, Ethics Commission's complete report on automated and connected driving, 2017 - <https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.html>

⁵⁸ Federal Government, The Federal Government's action plan on the report by the Ethics Commission on Automated and Connected Driving (Ethical rules for self-driving computers), 2017 - <https://www.bmvi.de/SharedDocs/EN/publications/action-plan-on-the-report-ethics-commission-acd.html>

2. The protection of individuals takes precedence over all other utilitarian considerations. The objective is to reduce the level of harm until it is completely prevented. The licensing of automated systems is not justifiable unless it promises to produce at least a diminution in harm compared with human driving, in other words a positive balance of risks.
3. The public sector is responsible for guaranteeing the safety of the automated and connected systems introduced and licensed in the public street environment. Driving systems thus need official licensing and monitoring. The guiding principle is the avoidance of accidents, although technologically unavoidable residual risks do not militate against the introduction of automated driving if the balance of risks is fundamentally positive.
4. The personal responsibility of individuals for taking decisions is an expression of a society centred on individual human beings, with their entitlement to personal development and their need for protection. The purpose of all governmental and political regulatory decisions is thus to promote the free development and the protection of individuals. In a free society, the way in which technology is statutorily fleshed out is such that a balance is struck between maximum personal freedom of choice in a general regime of development and the freedom of others and their safety.
5. Automated and connected technology should prevent accidents wherever this is practically possible. Based on the state of the art, the technology must be designed in such a way that critical situations do not arise in the first place. These include “dilemma situations”, in other words a situation in which an automated vehicle has to “decide” which of two evils, between which there can be no trade-off, it necessarily has to perform. In this context, the entire spectrum of technological options – for instance from limiting the scope of application to controllable traffic environments, vehicle sensors and braking performance, signals for persons at risk, right up to preventing hazards by means of “smart” road infrastructure – should be used and continuously evolved. The significant enhancement of road safety is the objective of development and regulation, starting with the design and programming of the vehicles such that they drive in a defensive and anticipatory manner, posing as little risk as possible to vulnerable road users.
6. The introduction of more highly automated driving systems, especially with the option of automated collision prevention, may be socially and ethically mandated if it can unlock existing potential for damage limitation. Conversely, a statutorily imposed obligation to use fully automated transport systems or the causation of practical inescapability is ethically questionable if it entails submission to technological imperatives (prohibition on degrading the subject to a mere network element).
7. In hazardous situations that prove to be unavoidable, despite all technological precautions being taken, the protection of human life enjoys top priority in a balancing of legally protected interests. Thus, within the constraints of what is technologically feasible, the systems must be programmed to accept damage to animals or property in a conflict if this means that personal injury can be prevented.
8. Genuine dilemmatic decisions, such as a decision between one human life and another, depend on the actual specific situation, incorporating “unpredictable” behaviour by parties affected. They can thus not be clearly standardized, nor can they be programmed such that they are ethically unquestionable. Technological systems must be designed to avoid accidents. However, they cannot be standardized to a complex or intuitive assessment of the impacts of an accident in such a way that they can replace or anticipate the decision of a responsible driver with the moral capacity to make correct judgements. It is true that a human driver would be acting unlawfully if he killed a person in an emergency to save the lives of one or more other persons, but he would not necessarily

be acting culpably. Such legal judgements, made in retrospect and taking special circumstances into account, cannot readily be transformed into abstract/general ex ante appraisals and thus also not into corresponding programming activities. For this reason, perhaps more than any other, it would be desirable for an independent public sector agency (for instance a Federal Bureau for the Investigation of Accidents Involving Automated Transport Systems or a Federal Office for Safety in Automated and Connected Transport) to systematically process the lessons learned.

9. In the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited. It is also prohibited to offset victims against one another. General programming to reduce the number of personal injuries may be justifiable. Those parties involved in the generation of mobility risks must not sacrifice non-involved parties.
10. In the case of automated and connected driving systems, the accountability that was previously the sole preserve of the individual shifts from the motorist to the manufacturers and operators of the technological systems and to the bodies responsible for taking infrastructure, policy and legal decisions. Statutory liability regimes and their fleshing out in the everyday decisions taken by the courts must sufficiently reflect this transition.
11. Liability for damage caused by activated automated driving systems is governed by the same principles as in other product liability. From this, it follows that manufacturers or operators are obliged to continuously optimize their systems and also to observe systems they have already delivered and to improve them where this is technologically possible and reasonable.
12. The public is entitled to be informed about new technologies and their deployment in a sufficiently differentiated manner. For the practical implementation of the principles developed here, guidance for the deployment and programming of automated vehicles should be derived in a form that is as transparent as possible, communicated in public and reviewed by a professionally suitable independent body.
13. It is not possible to state today whether, in the future, it will be possible and expedient to have the complete connectivity and central control of all motorvehicles within the context of a digital transport infrastructure, similar to that in the rail and air transport sectors. The complete connectivity and central control of all motorvehicles within the context of a digital transport infrastructure is ethically questionable if, and to the extent that, it is unable to safely rule out the total surveillance of road users and manipulation of vehicle control.
14. Automated driving is justifiable only to the extent to which conceivable attacks, in particular manipulation of the IT system or innate system weaknesses, do not result in such harm as to lastingly shatter people's confidence in road transport.
15. Permitted business models that avail themselves of the data that are generated by automated and connected driving and that are significant or insignificant to vehicle control come up against their limitations in the autonomy and data sovereignty of road users. It is the vehicle keepers and vehicle users who decide whether their vehicle data that are generated are to be forwarded and used. The voluntary nature of such data disclosure presupposes the existence of serious alternatives and practicability. Action should be taken at an early stage to counter a normative force of the factual, such as that prevailing in the case of data access by the operators of search engines or social networks.
16. It must be possible to clearly distinguish whether a driverless system is being used or whether a driver retains accountability with the option of overruling the system. In the case of non-driverless systems, the human-machine interface must be designed such that at any time it is clearly regulated and apparent on which side the individual

responsibilities lie, especially the responsibility for control. The distribution of responsibilities (and thus of accountability), for instance with regard to the time and access arrangements, should be documented and stored. This applies especially to the human-to-technology handover procedures. International standardization of the handover procedures and their documentation (logging) is to be sought in order to ensure the compatibility of the logging or documentation obligations as automotive and digital technologies increasingly cross national borders.

17. The software and technology in highly automated vehicles must be designed such that the need for an abrupt handover of control to the driver (“emergency”) is virtually obviated. To enable efficient, reliable and secure human-machine communication and prevent overload, the systems must adapt more to human communicative behaviour rather than requiring humans to enhance their adaptive capabilities.
18. Learning systems that are self-learning in vehicle operation and their connection to central scenario databases may be ethically allowed if, and to the extent that, they generate safety gains. Self-learning systems must not be deployed unless they meet the safety requirements regarding functions relevant to vehicle control and do not undermine the rules established here. It would appear advisable to hand over relevant scenarios to a central scenario catalogue at a neutral body in order to develop appropriate universal standards, including any acceptance tests.
19. In emergency situations, the vehicle must autonomously, i.e. without human assistance, enter into a “safe condition”. Harmonization, especially of the definition of a safe condition or of the handover routines, is desirable.
20. The proper use of automated systems should form part of people’s general digital education. The proper handling of automated driving systems should be taught in an appropriate manner during driving tuition and tested.

The Federal Government's Artificial Intelligence (AI) Strategy -2018

On November 2018, the Federal Cabinet adopted the Federal Government's Artificial Intelligence (AI) Strategy⁵⁹, on the basis of the proposal jointly submitted by the Federal Ministry of Education and Research, the Federal Ministry of Economy and Energy and the Federal Ministry of Labour and Social Affairs. The Strategy pursues the following three objectives:

1. making Germany and Europe global leaders on the development and use of AI technologies and securing Germany's competitiveness in the future,
2. safeguarding the responsible development and use of AI which serves the good of society, and
3. integrating AI in society in ethical, legal, cultural and institutional terms in the context of a broad societal dialogue and active political measures.

The strategy is based on a holistic approach, comprising twelve fields of action. It focuses on strengthening research in Germany and Europe, accelerating the transfer of research findings to businesses, promoting the availability of skilled workers and experts, shaping the structural change in enterprises and on the labour market, creating an environment for the ethical use of artificial intelligence, deepening European and international cooperation on AI issues and fostering the societal dialogue on the opportunities and the impact of artificial intelligence.

The Data Ethics Commission - 2018

⁵⁹Federal Government, Federal Government's Artificial Intelligence (AI) Strategy, 2018 - <https://www.bmwi.de/Redaktion/EN/Pressemitteilungen/2018/20181116-federal-government-adopts-artificial-intelligence-strategy.html>

To address the ethical issues of AI, as part of the actions in the 2018 Strategy, the federal government created the *Data Ethics Commission* ('Daten Ethik Kommission')⁶⁰. This Commission has the task of drawing on scientific and technical expertise to develop ethical guidelines for the protection of the individual, the preservation of social cohesion and the safeguarding and promotion of well-being in the information age. The Commission's work resulted in drafting two documents: "*Opinions of the Data Ethics Commission*".⁶¹ and "*Recommendations of the Data Ethics Commission for the Federal Government's Artificial Intelligence Strategy*".⁶². In the latter document, the Commission recommended that the government should:

- Uphold the ethical and legal principles based on our liberal democracy throughout the development and application of artificial intelligence.
- Promote the ability of individuals and society to understand and think critically in the information society.
- Promote and demand attention to ethical and legal principles in the entire AI development and application process, and the strategy should include this as an additional goal. The action areas of the strategy should be defined with this objective in mind.
- Include another area of action in the strategy focusing on creating appropriate framework conditions to promote the ability of individuals and society as a whole to understand and critically reflect on the information society.

Ethics and Innovation according to the president of Germany, F.W. Steinmeier - 2022

On 7 February 2022, the President of the Federal Republic of Germany *Frank-Walter Steinmeier* attended the presentation ceremony of the *Ethics of Digitalisation project*⁶³, a 2019 initiative of the Humboldt Institute for Internet and Society (HIIG), whose partners are the Leibniz-Institut für Medienforschung | Hans-Bredow-Institut, the Berkman Klein Center at Harvard University and the Digital Asia Hub. The international research project "*Ethics of Digitalisation - From Principles to Practices*" aims to develop innovative responses to the challenges in the conflict between ethics and digitalisation. The project's core is based on interdisciplinary scientific work on application- and practice-oriented issues to achieve high relevance and social impact results, promoting an active Exchange between science, policy and society and thus contributing to a global dialogue on the ethics of digitisation.

During the ceremony, held in Schloss Bellevue, President Steinmeier said the following⁶⁴: "Our world is becoming increasingly digital, and many people have seen over the past few years what that can mean for their own lives. Just last Thursday I was at a conference of works councils organised by a major German trade union. I witnessed their debate over the digitalisation of the world of work, and it reminded me of the origins of this project – of the need for an ethics of digitalisation to offer us support and guidance amid the digital transformation of business and science, of debate and democracy, of our society and our coexistence I firmly believe that, when it comes to topics such as artificial intelligence or the question of digital self-determination,

⁶⁰Data Ethics Commission -

https://www.bmj.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html

⁶¹Data Ethics Commission, "Opinion of the Data Ethics Commission", 2020 -

https://www.bmj.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.pdf?__blob=publicationFile&v=3

⁶²Data Ethics Commission, "Recommendations of the Data Ethics Commission for the Federal Government's Strategy on Artificial Intelligence", 2019 -

https://www.bmj.de/SharedDocs/Downloads/DE/Ministerium/ForschungUndWissenschaft/DEK_Empfehlungen_englisch.html;jsessionid=AD50394A578CC2BDBE32F8753B7CA7BC.1_cid297?nn=11678512

⁶³HIIG, Ethics of Digitalisation - <https://www.hiig.de/en/project/the-ethics-of-digitalisation/>

⁶⁴Bundespräsident, Federal President Frank-Walter Steinmeier at the presentation of the results of the project The Ethics of Digitalisation on 7 February 2022 at Schloss Bellevue, 2022 -

https://www.bundespraesident.de/SharedDocs/Downloads/DE/Reden/2022/02/220207-Ethik-der-Digitalisierung-Englisch2.pdf;jsessionid=D7F4EC80E50F29FEA3BEDC6FD946EBF4.1_cid393?__blob=publicationFile

when such major ethical questions remain unresolved, then we must always look beyond the confines of our national borders”.

Finally, President Steinmeier illustrated his approach to the relationship between ethics and innovation:

"What is very ... important for me - concluded the president - is that ethics and innovation cannot be a question of: either one or the other! Healthy competition for the best ideas, the smartest applications and the smartest products can only develop when the basic standards have been clearly defined"

Sitography

Canada

- CIFAR, *Pan-Canadian AI Strategy* - <https://cifar.ca/ai/>
- Statistics Canada, *Responsible use of automated decision systems in the federal government* - <https://www.statcan.gc.ca/en/data-science/network/automated-systems>
- Teresa Scassa, *Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada's Directive on Automated Decision-Making* (October 30, 2020). Forthcoming, (2021) 54:1 University of British Columbia Law Review [Scassa], online: <https://ssrn.com/abstract=3722192> at 6-7.
- Open Government - <https://open.canada.ca/data/en/dataset/748a97fb-6714-41ef-9fb8-637a0b8e0da1>

United States

- OSTP, *Preparing for the future of Artificial Intelligence*, 2016 https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
- OSTP, *National Artificial Intelligence Research and Development Strategic Plan*, 2016 https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf
- Executive Office of the President, *American AI Initiative*, 2019 <https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence>
- Executive Office of the President. (2020). *Guidance for regulation of Artificial Intelligence applications*. <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>
- US Congress, *National AI Initiative Act*, 2021 - <https://www.congress.gov/bill/116th-congress/house-bill/6216/text#toc-H06C2598E670D416096729065FCC48CF9>
- DLA Piper, *Landmark artificial intelligence legislation advances toward becoming law*, 2020 - <https://www.dlapiper.com/en/us/insights/publications/2020/12/landmark-artificial-intelligence-legislation-advances-toward-becoming-law/>
- Office of Science and Technology Policy. (2020). *President Trump's FY 2021 budget commits to double investments in key industries of the future* – The White House. <https://trumpwhitehouse.archives.gov/briefings-statements/president-trumps-fy-2021-budget-commits-double-investments-key-industries-future/>
- Roberts, H., Cowls, J., Hine, E. et al.: *Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US*. Sci Eng Ethics 27, 68 (2021) - <https://doi.org/10.1007/s11948-021-00340-7>

Russian Federation

- Corriere della Sera (2017) - https://www.corriere.it/tecnologia/economia-digitale/17_settembre_04/putin-sull-intelligenza-artificiale-chi-sviluppa-migliore-governo-mondo-musk-rilancia-l-allarme-c2a46c9c-916f-11e7-8332-148b1c29464d.shtml
- Stephanie Petrella, Chris Miller, and Benjamin Cooper, *Russia's Artificial Intelligence Strategy: The Role of State-Owned Firms*, November 2020 <https://sites.tufts.edu/hitachi/files/2021/02/1-s2.0-S0030438720300648-main.pdf>
- “Conference “Artificial Intelligence: Problems and Solutions – 2018”, Ministry of Defense of the Russian Federation, 2018 - <http://mil.ru/conferences/is-intellekt.htm>.
- Samuel Bendett, *Here's How the Russian Military Is Organizing to Develop AI*, Defense One, July 20, 2018, <https://www.defenseone.com/ideas/2018/07/russian-militarys-ai-development-roadmap/149900/>.
- *The President signed Executive Order On National Goals and Strategic Objectives of the Russian Federation through to 2024*, President of Russia, May 7, 2018 - <http://en.kremlin.ru/events/president/news/57425>
- *Rosatom summed up the results of tenders for the development of 'roadmaps' for end-to-end technologies*, D-Russia.ru, April 1, 2019 - <https://d-russia.ru/rosatom-podvelitgi-konkursov-na-razrabotku-dorozhnyh-kart-po-skvoznym-tehnologiyam.html>.
- *Artificial Intelligence Code of Ethics* - <https://a-ai.ru/wp-content/uploads/2021/10/Code-of-Ethics.pdf>
- TASS Russian News Agency, *First code of ethics of artificial intelligence signed in Russia*, 26 Oct. 2021 - <https://tass.com/economy/1354187>
- Forbes, *Tutte le paure del mondo: cosa temono i creatori del Codice Etico dell'Intelligenza Artificiale*, 1° novembre 2021 - <https://www.forbes.ru/tekhnologii/444503-vse-strahi-miracego-boatsa-sozdateli-kodeksa-etiki-iskusstvennogo-intellekta>
- Agenzia Sputnik, *Non nuocere: gli esperti hanno raccontato cosa cambierà dopo la comparsa del Codice Etico per l'Intelligenza Artificiale*, 1° novembre 2021 - <https://spbdnevnik.ru/news/2021-11-01/ne-navredi-eksperty-rasskazali-chto-izmenitsya-posle-poyavleniya-kodeksa-etiki-iskusstvennogo-intellekta>

European Union

- European Parliament, *Civil Law Rules on Robotics*, 2016 https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html
- European Commission, *EU Declaration on Cooperation on Artificial Intelligence*, 2018 - <https://ec.europa.eu/jrc/communities/en/node/1286/document/eu-declaration-cooperation-artificial-intelligence>
- High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines For Trustworthy AI*, 2019 - <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>
- European Commission, *Communication on Fostering a European approach to Artificial Intelligence*, 2021 - <https://digital-strategy.ec.europa.eu/news-redirect/709089>
- European Commission, *Coordinated Plan on Artificial Intelligence 2021 Review* - <https://digital-strategy.ec.europa.eu/news-redirect/709091>
- European Commission, *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* - <https://digital-strategy.ec.europa.eu/news-redirect/709090>
- Roberts, H., Cows, J., Hine, E. et al. *Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US*. *Sci Eng Ethics* 27, 68 (2021) - <https://doi.org/10.1007/s11948-021-00340-7>
- European Parliament and Council, *General Data Protection Regulation*, 2016 - <https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32016R0679>

- The IEEE *Initiative on Ethics of Autonomous and Intelligent Systems*. (2017). EthicallyAligned Design, v2. Retrieved September 18, 2018 from <https://ethicsinaction.ieee.org>
- Brattberg, E., Csernaton, R., & Rugova, V. (2020). *Europe and AI: Leading, lagging behind, or carving its own way?* Carnegie Endowment for International Peace. <https://carnegieendowment.org/2020/07/09/europe-and-ai-leading-lagging-behind-or-carving-its-own-way-pub-82236>
- European Commission. (2020b). *New EU financing instrument of up to €150 million to support European Artificial Intelligence companies*. <https://digital-strategy.ec.europa.eu/en/news/new-eu-financing-instrument-eu150-million-support-european-artificial-intelligence-companies>
- European Data Protection Supervisor. (2021). *Artificial Intelligence act: A welcomed initiative, but ban on remote biometric identification in public space is necessary*. https://edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative_e

Bibliography

Canada

- Treasury Board of Canada Secretariat, *Directive on Automated Decision-Making* (Ottawa: T.B.S., 5 February 2019)
- National Research Council Canada, *Advisory statement on human ethics in artificial intelligence and big data research* [Internet]. 2017
- Treasury Board of Canada Secretariat, *Responsible Artificial Intelligence in the Government of Canada*, Digital Disruption White Paper Series (10 April 2018)

United States

- Exec. Order No. 13859, § 84 Fed. Reg. 3967 (Feb. 11, 2019)
- Exec. Order No. 13563 § 76 Fed. Reg. 3821 (Jan. 21, 2011)
- National Artificial Intelligence Initiative Act of 2020, H.R. 6216, 116th Cong. (2020)

Russian Federation

- President of Russian Federation, Dec. 24, 2019, Defence Ministry Board meeting
- President of Russian Federation, Decree No. 204, Signed 07 May 2018, *On national objectives and strategic tasks of Russian Federation's development in the period up to 2024*
- AI Alliance Russia, Oct. 2021, *The Code of Ethics in the Field of Artificial Intelligence*

European Union

- European Commission, *Proposal for a Regulation of the European Parliament and for the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts*, COM(2021) 206 final, 2021
- European Commission, *White Paper on Artificial Intelligence - A European approach to excellence and trust*, COM(2020) 65 final, 2020.
- European Parliament resolution of 20 October 2020 with *Recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies*, 2020/2012(INL), (2021/C 404/04)
- European Parliament resolution of 16 February 2017 with *Recommendations to the Commission on Civil Law Rules on Robotics*, 2015/2103(INL), P8_TA(2017)0051

- European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Empty, COM(2021) 205 final, 2021
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on *The protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC* (General Data Protection Regulation), COM(2012)010 final, 2012

Nazioni Unite

UNESCO–IFAP (2021): *UGRA Memorandum. Tangible and Intangible Impact of Information and Communication in the Digital Age*, Final Document of the Third International Conference, Khanty-Mansiysk, 17–18 June 2021

UNITED NATIONS (2015): *Transforming our World. The 2030 Agenda for Sustainable Development*, A/RES/70/1, New York, September 2015

Website: sustainabledevelopment.un.org

UNITED NATIONS (2019): *The Future is Now: Science for Achieving Sustainable Development*, Global Sustainable Development Report (GSDR) 2019 New York, 11.9.2019

Website: <https://sustainabledevelopment.un.org/globalreport/2019>

UNITED NATIONS (ed.) (2019): *The Age of Digital Interdependence*. Report of the UN Secretary-General’s High-level Panel on Digital Cooperation, 2019, <https://eucyberdirect.eu/wp-content/uploads/2019/10/un-highlevel-panel-digital-cooperation-2019.pdf>.

UNITED NATIONS, António Guterres (2022): *2022 Priorities*, UN General Assembly, UNIFEED, New York City, 21 January 2022

References

EURISPES INSTITUTE

International Department

Rome (Italy)

Marco Ricceri , Ida Nicotera

Email: eurispes.intl-dept@libero.it

Web: www.eurispes.eu