AGID | Agenzia per l'Italia Digitale

Draft guidelines for the adoption of AI in public administration

Pursuant to the Prime Ministerial Decree of 12 January 2024, 'Three-year plan for information technology in public administration 2024–2026'

Draft version 4.0 of 9 March 2026.

Working group

This document has been drawn up within the Consultation Table of the three-year plan for information technology in public administration. The following took part in the work:

- AgID – Agency for Digital Italy
- ACN – National Cybersecurity Agency
- ANAC – National Anti-Corruption Authority
- ANCI – National Association of Italian Municipalities
- Conference of Regions and Autonomous Provinces
- Consip S.p.A.
- Department of Public Administration of the Prime Minister's Office
- Department for Digital Transformation of the Prime Minister's Office
- INAIL – National Institute for Insurance against Accidents at Work
- INPS - National Social Security Institute
- ISTAT – Italian National Institute of Statistics
- IPZS – State Printing Office and Mint
- Ministry of the Economy and Finance
- Ministry of Enterprises and Made in Italy
- PagoPA S.p.A.
- Union of Italian Provinces

Table of contents

Introduction

In the context of the use of Artificial Intelligence (AI) systems by public administrations (PA), including for the purposes of Article 2(1) of Legislative Decree No 82 of 7 March 2005 laying down the Digital Administration Code ('the DAC'), the Guidelines for the adoption, procurement and development of artificial intelligence systems in public administration are laid down in the Prime Ministerial Decree of 12 January 2024, 'Three-year plan for information technology in public administration 2024–2026'.

The Guidelines for the adoption, procurement and development of artificial intelligence systems in public administration are issued following the procedure laid down in Article 71 of the DAC and in compliance with the Decision No 160 of the Agency for Digital Italy ('AgID') of 17 May 2018 laying down regulations for the adoption of Guidelines for the implementation of the Digital Administration Code.

The Guidelines form part of the existing regulatory framework on artificial intelligence at both EU and national level, with particular reference to Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) and Law No 132 of 23 September 2025 on provisions and powers delegated to the Government in relation to artificial intelligence.

With reference to the processing carried out for the purposes of care and scientific research, these Guidelines refer to the specific provisions laid down in the aforementioned Regulation (EU) 2024/1689, in 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), in Legislative Decree No 196 of 30 June 2003, as amended, containing the Personal Data Protection Code, in the aforementioned Law No 132 of 23 September 2025 and in the specific sectoral regulations.

1. Scope of application

1.1. Subjective scope

These Guidelines for the Adoption of Artificial Intelligence in Public Administration ('the Guidelines') are intended for the entities referred to in Article 2(2) of the DAC. These entities are referred to in this document by the acronym 'PA'.

1.2. Objective scope

These Guidelines concern the procedures for adopting artificial intelligence systems, with particular reference to regulatory compliance and organisational impact.

They apply to all components of technological applications and infrastructure that utilise artificial intelligence technologies, whether as integrated components or to support core functionalities.

2. References and acronyms

2.1. Document reading notes

These Guidelines adopt the key terms 'MUST', 'MUST NOT', 'SHOULD', 'MAY' and 'CANNOT', as defined in the ISO/IEC Directives, Part 2 *'Principles and rules for the structure and drafting of ISO and IEC documents'*. The interpretation of these terms within the framework of the Guidelines is described below.

- MUST indicates a mandatory requirement;
- MUST NOT or CANNOT indicate an absolute prohibition;
- SHOULD or SHOULD NOT indicate that the implications must be understood and carefully weighed before choosing alternative approaches;
- MAY or the adjective OPTIONAL indicate that the reader may choose whether or not to apply the specification.

These Guidelines refer to concepts and principles contained in technical standards defined at national, EU and international level. These standards are not binding, unless otherwise expressly stated in the text.

2.2. Structure

Given the speed of innovation, the Guidelines must ensure constant adaptation to the changes imposed by the continuing digital revolution. Hence the decision to supplement the Guidelines with 'Tools', i.e. documents designed to support the practical implementation of the Guidelines.

The updated list of tools in the AI Guidelines is available at: <<to be determined at the time of publication>>.

The Tools may be periodically updated to take account of regulatory and technological developments and emerging best practices.

2.3. Regulatory framework

Below are the main acts used as the regulatory framework for these Guidelines.

[CFREU]	Charter of Fundamental Rights of the European Union C2012/326/02
[AI Act]	Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).
[Data Act]	Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act).
[DGA]	Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).
[CRA]	Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act).
[Regulation (EU) 2018/1725]	Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC.
[GDPR]	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).
[Regulation (EU)	Regulation (EU) No 2012/1025 of the European Parliament

2012/1025]	and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council.
[Open Data Directive]	Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast).
[NIS2]	Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS2 Directive).
[CER]	Directive (EU) 2022/2557of the European Parliament and of the Council of 14 December 2022 on the resilience of critical entities and repealing Council Directive 2008/114/EC (Critical Entities Resilience Directive).
[Directive (EU) 2016/680]	Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.
[Regulation (EU) 2008/765]	Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93 (New Legislative Framework).
(Decision No	Decision No 768/2008/EC of the European Parliament and of

768/2008/EC]	the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC (New Legislative Framework).
[Regulation (EU) 2019/1020]	Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011 (New Legislative Framework).
[Products Directive]	Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC.[1]
[AI Law]	Law No 132 of 23 September 2025 on provisions and powers delegated to the Government in relation to artificial intelligence.
[Law No 90/2024]	Law No 90 of 28 June 2024 laying down provisions on strengthening national cybersecurity and cybercrime.
[Decree-Law No 135/2018]	Decree-Law No 135 of 14 December 2018 laying down urgent provisions on support and simplification for businesses and public administration, converted into law, with amendments, by Article 1(1) of Law No 12 of 11 February 2019.
[Accessibility]	Law No 4 of 9 January 2004, as amended, on provisions to facilitate and simplify the access of users and, in particular, persons with disabilities to IT tools.
[Privacy Code]	Legislative Decree No 196 of 30 June 2003, as amended, concerning the Data Protection Code, laying down provisions for the adaptation of national law to 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.
[DAC]	Legislative Decree No 82 of 7 March 2005, as amended, on

[1] Directive (EU) 2024/2853 will enter into force on and must be transposed into national law by 9 December 2026.

	the Digital Administration Code.
[Legislative Decree No 36/2006]	Legislative Decree No 36 of 24 January 2006 implementing Directive (EU) 2019/1024 on open data and the re-use of public sector information which repealed Directive 2003/98/EC.
[Public Procurement Code]	Legislative Decree No 36 of 31 March 2023, as amended, containing the Public Contracts Code implementing Article 1 of Law No 78 of 21 June 2022, granting the Government delegated powers in the field of public contracts.
[Legislative Decree No 200/2021]	Legislative Decree No 200 of 8 November 2021 implementing Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast).
[Legislative Decree No 82/2022]	Legislative Decree No 82 of 27 May 2022 implementing Directive (EU) 2019/882 (CER) of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services.
[Legislative Decree No 134/2024]	Legislative Decree No 134 of 4 September 2024 implementing Directive (EU) 2022/2557 of the European Parliament and of the Council of 14 December 2022 on the resilience of critical entities and repealing Council Directive 2008/114/EC.
[Legislative Decree No 138/2024]	Legislative Decree No 138 of 4 September 2024 transposing Directive (EU) 2022/2555 (NIS2) on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148.
[Presidential Decree No 81/2022]	Presidential Decree No 81 of 24 June 2022. Regulation on the identification of requirements relating to plans absorbed by the integrated plan of activities and organisation (IPAO).
[Three-Year Plan]	Prime Ministerial Decree of 12 January 2024 setting out the three-year plan for IT in public administration 2024–2026

[TYP updated 2025]	Prime Ministerial Decree of 3 December 2024 on the 2025 update to the Three-Year Plan 2024–2026
[Digital Decade]	European Declaration on Digital Rights and Principles for the Digital Decade (2023/C 23/01)
[AI Strategy]	Italian Artificial Intelligence Strategy 2024–2026.

2.4. Reference Guidelines

The following are the guidelines issued by the AGID pursuant to Article 71 of the DAC and other regulatory documentation, which are also referred to indirectly in this document. The AgID guidelines are available on the institutional website at https://www.agid.gov.it/it/linee-guida, where relevant updates, as a result of technological developments or the need to adapt to the reference legislation, are also published.

[AI GL]	Guidelines for the adoption, procurement and development of artificial intelligence systems in public administration <<link to be provided upon publication>>
[OPEN DATA GL]	Open Data Guidelines.
[REUSE GL]	Guidelines for the acquisition and reuse of software for public administrations.
[ACCESS GL]	Guidelines on the accessibility of IT tools
[DES GL]	Design guidelines for PA's digital websites and services
[COMP DOC GL]	Guidelines on the formation, management and preservation of computerised documents
[PDND INTER GL]	Guidelines on the technological infrastructure of the National Digital Data Platform for the interoperability of database information systems
[SEC INTER GL]	Technological guidelines and standards for the security of interoperability through information system APIs
[TECH INTER GL]	Guidelines on technical interoperability of public administrations

[eESPD TS]	Technical specifications for the definition of the 'eDGUE-IT' Italian electronic ESPD
[CLOUD REG]	Single Regulation No 21007 on cloud infrastructure and services for public administrations, adopted by the ACN on 27 June 2024.
[AI_ACT_PROHIB]	C(2025) 884 Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act).

2.5. Acronyms

The acronyms used in these **Guidelines** are set out below.

[PA]	All entities referred to in Article 2(2) of the DAC.
[DPIA]	Data protection impact assessment pursuant to Article 35 of the GDPR and Article 27 of Directive No 2016/680.
[FRIA]	Fundamental rights impact assessment for high-risk AI systems pursuant to Article 27 AI Act.
[GPAI]	General-purpose AI system.
[KPI]	Key Performance Indicator.
[NLF]	New Legislative Framework: the European Union's regulatory framework for the harmonisation of product legislation.
[IPAO]	Integrated plan of activities and organisation referred to in Presidential Decree No 81/2022.
[NRRP]	National Recovery and Resilience Plan.
[POC]	Proof of Concept.
[QoS]	Quality of Service: an indication of the parameters used to characterise the quality of e-services.
[AIM]	Artificial Intelligence Manager.
[DPO]	Data Protection Officer.
[DTM]	Digital Transition Manager.
[SDG]	Sustainable Development Goals.

[SLA] Service Level Agreement: agreement on the level of service resulting from the bargaining between the provider and the user.

[SLI] Service-Level Indicator: a metric to measure the efficiency of the services identified by the provider.

[SLO] Service-Level Objective: the objectives of the SLOs for services defined by the provider.

[TRL] Technology Readiness Level.

[DTO] Digital Transition Office.

2.6. Terms and definitions

For the terms and definitions used in this document, please refer to the 'Terms and Definitions' section of the AgID Guidelines on Artificial Intelligence.

The tool brings together the main technical, regulatory and operational concepts related to AI. The terms and definitions are aligned, where applicable, with EU legislation and relevant international standards.

3. Artificial Intelligence

Artificial Intelligence (AI) is a set of technologies capable of transforming and enhancing economic and social activities, thereby improving decision-making processes, operational efficiency and the quality of services offered to organisations and individuals.

The term 'AI system' refers to *'a machine-based system[2] designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments'*[3].

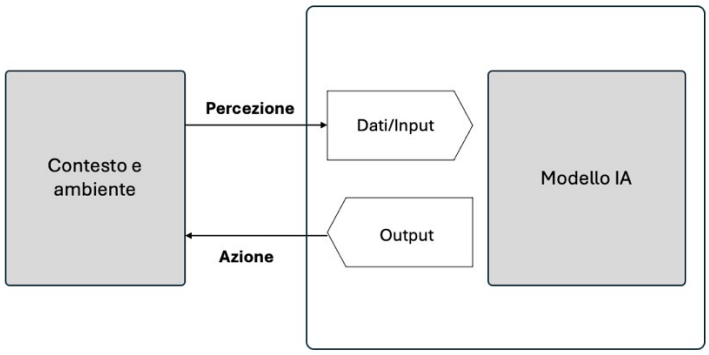


Figure 1 – AI system[4]

Contesto e ambiente	Environment (context)
Percezione	Perceive
Azione	Influence
Dati/Input	Data & input
Output	Outputs
Modello IA	AI model

A distinctive feature of AI systems is the inferential capability [5] of their core component, the AI model[6]; this component, trained on large amounts of data, enables the system to generate outputs such as predictions, content,

[2] The Italian version of the AI Act translates the English term 'machine-based' as 'automatizzato' ('automated').

[3] AI Act, Article 3(1). For the purposes of the AI Act, the 'Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)' should also be taken into account.
https://ec.europa.eu/newsroom/dae/redirection/document/112455

[4] The image is taken from the OECD AI Principles overview: https://oecd.ai/en/ai-principles. The definition of AI system in the AI Act was aligned with the OECD definition during the trilogue negotiations on the AI Act.

[5] Inference is the process of deriving new information, conclusions or predictions from existing data, rules or knowledge. In the context of AI, it refers to the ability of a model or system to process input data to generate significant outputs, such as classifications, recommendations, decisions or predictions. See: ISO/IEC 22989:2022 - Artificial Intelligence - Concepts and terminology https://www.iso.org/standard/74296.html.

recommendations or decisions based on the input data. This capacity goes beyond the simple or mechanical processing of data, enabling the system to:

- learn from data: use machine-learning techniques to identify patterns in data, gain knowledge and achieve specific goals;
- reason and deduce: apply approaches based on logic and knowledge, which allow the system to draw inferences using predefined rules, symbolic representations or codified knowledge;
- model: create abstract representations of the problem or context to support complex decision-making processes or provide innovative solutions.

Inferential capability enables AI systems to go beyond the static data processing typical of traditional software, adapting and improving over time based on the outputs achieved and operating conditions.

Unlike deterministic systems, AI systems – particularly those based on machine-learning models – produce outputs that reflect a statistical and probabilistic logic: the predictions, recommendations or decisions generated are derived from inferences learned from past data and are therefore subject to a degree of uncertainty.

This characteristic requires particular attention when assessing the reliability and robustness of outputs, especially in high-impact contexts, where the effects of automated decisions can affect fundamental rights, social fairness or access to essential public services.

The objectives of an AI system may differ from the specific purpose for which it is used in a given context; therefore, an AI system may be applied differently depending on the use case or operational context.

AI systems can operate with varying levels of autonomy, ranging from direct human control to independent operation.

Adaptability is a key feature of AI systems: thanks to machine learning, AI systems can modify their behaviour during use, thereby improving their performance over time.

[6] ISO/IEC 22989:2022 defines a model as 'a physical, mathematical or otherwise logical representation of a system, entity, phenomenon, process or data'. AI models include, among others, statistical models and various types of input–output functions (such as decision trees and neural networks).

The term 'general-purpose AI model' (GPAI) refers to: '*an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market*'[7].

Article 51 of the AI Act introduces a specific category of generative AI models (GPAI) classified as having high systemic risk, characterised by 'high impact capabilities' or a significant impact on the internal market. These models are subject to a specific regulatory framework set out in the AI Act.

Article 56 of the AI Act provides that the European Commission shall draw up a code of practice, to be adopted on a voluntary basis by providers of GPAI models posing a systemic risk, pending the submission of a formal request for standardisation to the EU standardisation bodies for the definition of specific standards for such models (see Article 40 of the AI Act).

AI systems may be subject to biases[8] or prejudices that may manifest themselves in the outputs or behaviours of an AI system due to several factors, such as:

- data bias: when the data used to train or test an AI model reflects prejudices, omissions or biased representations of reality, thereby adversely affecting the fairness and accuracy of the system;
- algorithmic bias: when the technical choices or assumptions implemented in the AI models or algorithms used introduce bias into the outputs;
- human bias: when implicit or explicit prejudices of designers or stakeholders are reflected in the design, training or application of the AI system.

Bias can lead to unfair treatment, unreliable outcomes or discrimination against specific individuals or groups. Recognising, mitigating and monitoring biases is essential to ensure that AI systems operate in a transparent and fair manner.

[7] AI Act, Article 3(63).

[8] ISO/IEC TR 24027 Artificial intelligence – Bias in AI systems and AI aided decision making https://www.iso.org/standard/77607.html

Article 10 of the AI Act provides, for high-risk systems, specific data governance requirements aimed at preventing bias. For further information on data governance, please refer to Chapter Error: Reference source not found.

3.1. Life cycle of an AI system

These Guidelines adopt as a reference model the life cycle of an AI system defined by the OECD. The stages that make up this life cycle, from initial planning through to decommissioning, are described below.

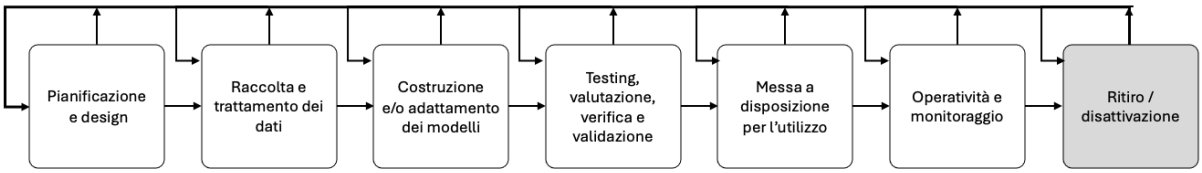


Figure 2 – The life cycle of an AI system[9]

Pianificazione e design	Planning & design
Raccolta e trattamento dei dati	Collect and process data
Costruzione e/o adattamento dei modelli	Build and/or adapt model(s)
Testing, valutazione, verifica e validazione	Test, evaluate, verify & validate
Messa a disposizione per l'utilizzo	Make available for use / deploy
Operatività e monitoraggio	Operate and monitor
Ritiro/disattivazione	Retire / decommission

- **Plan and design:** the organisation identifies the objective it intends to achieve, through the AI system, including the context in which the system will have to operate and the required data. This phase involves identifying the needs and requirements of stakeholders, translating these into measurable technical requirements and designing the system architecture and design (see Guidelines for AI Development).
- **Collect and process data:** data (both structured and unstructured) and the corresponding metadata are collected from various sources, analysed to determine whether they conform to statistical distributions in order to estimate the relevant parameters and then cleaned, enriched and transformed as required. These operations must follow the guidelines set out in the chapter on data management and quality (see Chapter Error: Reference source not found). The most relevant characteristics of the dataset are identified and selected, discarding any that are not of interest, with the aim of reducing the size of the dataset.

[9] OECD AI Principles overview, available at: https://oecd.ai/en/ai-principles.

- **Build and/or adapt model(s):** the organisation develops or selects an AI model that is suitable for achieving the system's objective and for the available data. Choosing the model also involves choosing the learning algorithms[10] with which the training will be carried out. The AI model is then trained on the basis of the data and the identified learning algorithm. The model tuning phase is next, during which the model is fine-tuned by adjusting the hyperparameters based[11] on a validation dataset. Alternatively, the organisation may opt for a 'transfer of learning' approach, acquiring a pre-trained and pre-configured AI model to use as a starting point for further training. This approach can be chosen when the organisation does not have sufficient data for the training.
- **Test, evaluate, verify & validate:** the AI system undergoes a series of rigorous tests to ensure it meets the requirements and expectations of stakeholders. During this phase, any anomalies are identified and resolved, while checks are carried out to ensure that the model operates correctly in the intended context and meets compliance and quality criteria.
- **Make available for use / deploy**: the trained AI system is made available to users and integrated into operational processes, ensuring compatibility with other systems and accessibility for all intended users. This phase includes making the AI system available for reuse.
- **Operate and monitor:** the AI system and input data are continuously monitored to detect any changes in operational patterns, data anomalies or the emergence of biases that could compromise the reliability of decisions. The organisation maintains, retrains and optimises the model, constantly assessing whether the objectives that were set have been met and assessing its operational effectiveness.
- **Retire / decommission:** when the AI system is no longer needed, it is retired or deactivated in a controlled manner. This phase includes securely managing data, documenting final operations and planning any replacements or alternatives.

[10] The three most common types of algorithm are supervised, unsupervised and reinforcement learning.

[11] Parameters that allow you to control the model's training process.

3.2. The roles in the AI value chain

The AI value chain involves different actors playing key roles in the development, deployment and use of AI technologies. These Guidelines adopt the definitions of provider and deployer provided by the AI Act[12] in the general context of AI systems. For a description of the specific obligations applicable to providers, deployers and other roles identified by the AI Act, such as importers and distributors, for high-risk AI systems, please refer to Chapter Error: Reference source not found.

- Provider: '*a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge*';
- Deployer: '*a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity*'.

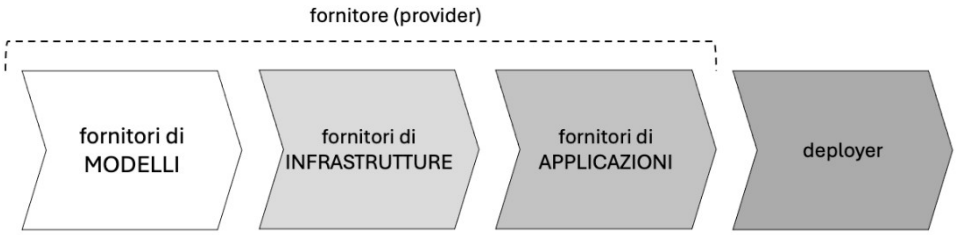


Figure 3 – The AI value chain

fornitore (provider)	provider
fornitore di MODELLI	provider of MODELS
fornitore di INFRASTRUCTURE	provider of INFRASTRUCTURE
fornitore di APPLICAZIONI	provider of APPLICATIONS
deployer	deployer

The AI value chain depicted in Figure 3 highlights four main levels:

- **Providers of models**

[12] AI Act, Article 3(3) and (4).

Entities that develop and/or provide AI models, specifically general-purpose AI (GPAI) models.

- **Providers of infrastructure**
  Entities that develop and/or provide AI infrastructure components such as data management, networking, hardware, cloud services and MLOps platforms.
- **Providers of applications**
  Entities that develop and/or provide applications for specific purposes by adapting AI and GPAI models. These actors have in-depth knowledge of the sectors and needs of user organisations, combining AI capabilities with domain expertise to customise solutions.
- ***Deployer***
  Organisations that use AI systems to meet operational or strategic needs. Deployers can integrate these systems into their own processes, customise them or adapt them (for example, by fine-tuning them using proprietary data), with the aim of optimising performance according to the application context. For the purposes of the AI Act, if a deployer significantly modifies the AI system or uses it under its own name or trademark, it assumes the responsibilities of a provider (see Article 25 of the AI Act).

3.3. Classification of AI systems based on risk

AI offers benefits in key sectors such as healthcare, education, transport, energy and public administration, but it may pose risks to public interests and fundamental rights protected by EU law. These Guidelines adopt the risk classification defined by the AI Act:

- **Prohibited AI systems**[13]*:* such systems pose an unacceptable risk and are therefore prohibited. Article 5 of the AI Act provides a description of the AI systems which the AI Act prohibits from being placed on the market as they may cause people significant harm. Examples include: AI systems that deploy subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with

[13] The obligations relating to the prohibited uses of AI under Article 5 of the AI Act entered into force on 2 February 2025. The European Commission has published 'Guidelines on artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)'.
https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

the objective of distorting a person's behaviour; AI systems that incorporate social scoring, resulting in biased or unfavourable treatment of certain groups of individuals.

- **High-risk AI systems:** systems that pose significant risks. In the event of a malfunction, such systems pose a significant potential for harm, with serious implications for public interests and fundamental rights. In order to classify a system as high-risk, public administrations MUST refer to the classification system set out in the AI Act (see, in particular, Article 6 and Annex III), including in relation to any possible derogations (see Article 6(3)). The European Commission is empowered to adopt delegated acts to add to or amend the use cases for high-risk AI systems set out in Annex III to the AI Act.
- **Limited-risk AI systems**: these are systems that pose fewer risks than the systems described above, and are subject to the transparency obligations set out in Article 50 of the AI Act. In the event of a malfunction, these systems are characterised by a limited potential for harm (e.g. generic chatbots), with a limited impact on public interests and fundamental rights.
- **AI systems with minimal or no risk**: these are systems with negligible risk. In the event of a malfunction, the resulting harm is very limited, as it is likely to have a negligible impact on public interests and fundamental rights.

For high-risk AI systems, in addition to the obligations set out in the AI Act, in order to ensure a consistently high level of protection of public interests in the areas of health, safety and fundamental rights, public administrations MUST:

- respect the principles set out in the **Charter of Fundamental Rights of the EU** and in the **European Convention on Human Rights**[14] by assessing the impact of AI systems on the rights protected therein;
- consider the **European Declaration on Digital Rights and Principles**[15] for the *Digital Decade*, ensuring the alignment of AI applications with European values;

[14] European Convention for the Protection of Human Rights and Fundamental Freedoms https://www.echr.coe.int/documents/d/echr/convention_ita

[15] European Declaration on Digital Rights and Principles https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:JOC_2023_023_R_0001.

- follow the **Ethics guidelines for trustworthy AI** of the HLEG on AI[16] by promoting transparency, fairness and accountability in the use of AI.

Chapter V of the AI Act defines a specific regulation for **general-purpose AI models with systemic risk**, which may have actual or reasonably foreseeable negative effects on public health, safety, fundamental rights or society as a whole.

3.4. Principles for the adoption of AI in public administrations

Public administrations shall adopt AI systems with a view to achieving, as a priority but not exclusively, the following objectives:

- to automate simple, repetitive tasks involving information research and analysis, freeing up time for higher-value activities;
- to increase predictive capabilities by improving data-driven decision-making;
- to support user-centred personalisation of services, thereby enhancing the effectiveness, efficiency and timeliness of public services, including through proactive measures;
- to promote innovation in public services and administrative processes.

Public administrations shall adopt AI systems in accordance with the following principles:

[16] Ethics guidelines for trustworthy AI https://data.europa.eu/doi/10.2759/640340

Compliance and governance

P.1 **Regulatory compliance.** Public administrations shall adopt AI systems in full compliance with national and EU regulations, ensuring adherence to the current legislative framework. Public administrations shall monitor regulatory developments and constantly update their systems to ensure compliance with current regulations.

P.2 **Respect for the EU's fundamental values.** Public administrations shall adopt AI systems in accordance with the principles set out in the Charter of Fundamental Rights of the European Union (see Articles 1 and 2 of the AI Act).

P.3 **Risk management.** Public administrations shall adopt appropriate risk management policies by carrying out a thorough analysis of the risks associated with the use of AI systems in order to prevent violations of fundamental rights or other significant harm.

P.4 **Personal data protection.** Public administrations shall adopt AI systems in compliance with current EU and national data protection regulations, ensuring high data quality and integrity (see Chapter 9.).

Ethics and inclusion

P.5 **Responsibility.** Public administrations shall adopt AI as a tool to support human activities, in the knowledge that the ultimate responsibility for decisions made by AI systems – whether automatically or under supervision – remains with the public administration. Public administrations shall clearly identify the responsibilities of all parties involved in the life cycles of AI systems.

P.6 **Accessibility, inclusiveness, non-discrimination.** Public administrations shall ensure fair treatment for all individuals and groups involved in the adoption of AI, promoting equal access, gender equality and cultural diversity. Public administrations shall take preventive measures to avoid the reproduction or amplification of biases present in society.

P.7 **Transparency, explainability and documentation.** Public administrations shall adopt AI systems while ensuring that their decisions and operations are transparent and comprehensible, in accordance with the principle of maintaining proportionality between transparency and effectiveness. Public administrations shall ensure that AI systems are developed and deployed in

such a way as to enable an adequate level of understanding of how they work. Public administrations shall ensure that AI systems are properly documented throughout their life cycles. Articles 11 and 13 of the AI Act reinforce this principle for high-risk systems.

P.8 **Transparency and information.** Public administrations shall inform users about how to interact with AI systems, making them aware of the capabilities and limitations of such systems. Article 50 of the AI Act reinforces this principle for systems intended to interact directly with natural persons (see Chapter 7.).

The quality and reliability of AI systems

P.9 **Data quality.** Public administrations shall adopt AI systems while ensuring that data is managed in an ethical, transparent and compliant manner. This includes using high-quality data, establishing clear policies governing access to, the use of and the retention of data, and implementing technical and organisational measures to safeguard the integrity and security of institutional data, ensuring that data is managed in a sustainable and responsible manner. Article 10 of the AI Act reinforces this principle for high-risk systems (see Chapter Error: Reference source not found).

P.10 **Accuracy**. Public administrations shall ensure that the outputs and decisions produced by artificial intelligence systems are accurate and consistent with the objectives that were set. To this end, measures must be put in place to continuously monitor performance so that any errors or deviations may be identified and corrected promptly. Article 15 of the AI Act reinforces this principle for high-risk systems.

P.11 **Robustness**. Public administrations shall ensure the robustness of AI systems, guaranteeing that they are able to operate correctly even under adverse conditions or in the event of a failure. Robustness refers to the ability to maintain the desired performance despite unforeseen events, data errors or technical issues. Article 15 of the AI Act reinforces this principle for high-risk systems.

P.12 **Cybersecurity.** Public administrations shall ensure the cybersecurity of AI systems, preventing attempts by third parties to tamper with, compromise or misuse them. This includes taking appropriate protective measures to safeguard the integrity, availability and confidentiality of the data and processes managed by AI systems. Article 15 of the AI Act reinforces this principle for high-risk systems.

P.13 **Human oversight.** Public administrations shall ensure an adequate level of human oversight of AI systems. For the purposes of human oversight, public administrations shall ensure that AI systems are designed and implemented in such a way as to allow them to be checked, corrected or replaced by human staff. Article 14 of the AI Act reinforces this principle for high-risk systems.

P.14 **Logging.** Public administrations shall adopt AI systems equipped with the appropriate logging mechanisms needed to track and retain records of

operations carried out over time. Article 12 of the AI Act reinforces this principle for high-risk systems.

P.15 Adoption of technical standards. Public administrations shall take due account of the technical standards established at national, EU and international level in order to ensure the interoperability, maintainability, security and compliance of AI systems with current legislation.

Innovation and sustainability

P.16 Efficiency and quality of services. Public administrations shall adopt AI to boost operational efficiency and improve the quality of services provided to people and businesses. Public administrations shall use AI to automate repetitive tasks related to institutional services and the operation of the administrative apparatus, encouraging the implementation of innovative and proactive solutions that simplify access to services and overall efficiency.

P.17 Innovation and continuous improvement. Public administrations shall adopt an approach to AI that focuses on innovation and continuous improvement, with a view to optimising administrative processes and the quality of services provided to people and businesses. Public administrations shall maintain collaborative relationships with other public administrations, research bodies, universities and businesses, actively involving stakeholders to test and integrate AI-based technological solutions, thereby creating an ecosystem conducive to innovation and the adoption of digital technologies. Public administrations shall consider the use of open-source AI systems and models, assessing their potential in terms of transparency, reuse, interoperability, technological autonomy and economic sustainability. Public administrations shall promote forms of collaboration and cooperation – including permanent arrangements or networks – between administrations, with a view to sharing expertise, infrastructure, data, technological solutions, development capabilities and procurement capacities.

P.18 Environmental sustainability. Public administrations shall adopt AI systems as part of a sustainable approach that prioritises environmental protection and is in line with the principles of energy sustainability.

Training and organisation

P.19 Training and skills development. Public administrations shall invest in training their staff to ensure they have the necessary skills to use, manage

and develop AI systems effectively and responsibly. Public administrations shall promote digital literacy initiatives aimed at people and businesses in order to foster a widespread and conscious understanding of the opportunities and implications of AI, ensuring that the adoption of AI in public services takes place in an inclusive and informed manner.

P.20 Strengthening organisation and infrastructure. When adopting AI systems, public administrations shall optimise their organisational structure and technological infrastructure – including through pooled or shared arrangements – to support the digital transformation process, strengthen control over digital systems and processes and improve the administration's resilience and operational autonomy, while ensuring adequate levels of security.

For AI systems classified as high risk, the principles set out above are reinforced by specific requirements laid down in the AI Act (see Chapter 5.).

4. Organisational model for the adoption of AI

Public administrations MUST adopt a structured set of processes, policies, resources and tools to govern, implement, monitor and improve the use of AI systems throughout their life cycles.

Public administrations MUST implement an organisational model for the adoption of AI that is capable of responding promptly to changes in the regulatory and technological landscape.

These Guidelines propose an AI adoption model based on the Plan-Do-Check-Act (PDCA) continuous improvement cycle and some management practices defined by the ISO/IEC 42001:2023 standard.[17] Figure 4 summarises the model, which is discussed in more detail in the following sections.

Public administrations MAY adopt this model, adapting it to their own characteristics, needs and responsibilities, with particular reference to their role as a provider or *deployer* of AI systems. In particular, public administrations MAY, when adopting AI systems with limited, minimal or no risk, combine the stages of the model, adapting them to the low complexity of the adoption project.

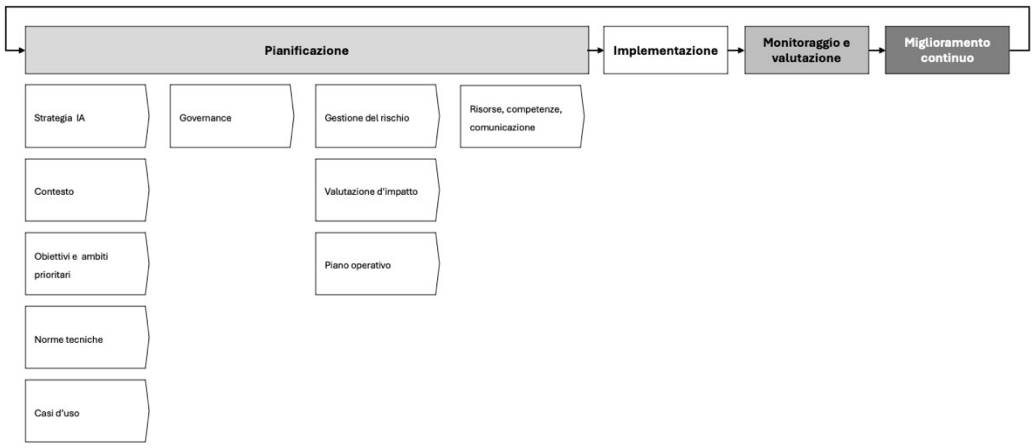


Figure 4. Illustration of the stages of the AI adoption model.

Pianificazione	**Planning**
Implementazione	**Implementation**
Monitoraggio e valutazione	**Monitoring and assessment**
Miglioramento continuo	**Continuous improvement**
Strategia IA	AI strategy
Contesto	Background
Obiettivi e ambiti prioritari	Objectives and priority areas
Norme tecniche	Technical regulations
Casi d'uso	Use cases
Governance	Governance
Gestione del rischio	Risk management

[17] ISO/IEC 42001:2023 Information technology - Artificial intelligence - Management system https://www.iso.org/standard/81230.html

Valutazione d'impatto	Impact assessment
Piano operativo	Operational plan
Risorse, competenze comunicazione	Resources, communication skills

4.1. AI strategy

Public administrations MUST develop an AI strategy that makes sense for their specific context (see Section 4.2.) and is aligned with their mission in terms of administrative function and specific operational areas.

Public administrations MAY draw up a joint networking strategy based on their specific characteristics and type (e.g. local authorities or universities).

The strategy MUST set out the objectives for the use of AI (see Section 4.3.) and the actions required to achieve them, promoting a shared and collaborative approach involving all departments of the public administration.

The strategy MUST include measures to ensure transparency regarding how the public administration uses AI systems.

Public administrations shall include the following actions in their AI strategies:

- improve the quality of data, including databases for documents and purchases (see Chapter 9.);
- improve the skills of internal staff on the subject of AI and personal data protection (see Chapter Error: Reference source not found);
- identify the use cases that are likely to deliver the greatest benefits (see Section 4.5.);
- assess the experiences and trials already carried out by other PAs, taking into account the opportunities for reuse (see Article 69 DAC);
- provide for the use of open-source AI systems and models;
- launch trials, including in partnership with other public administrations, starting with less complex use cases (quick wins);
- establish partnerships with other public administrations for running trials and subsequently adopting, procuring and developing AI systems on a collaborative basis.

Public administrations MUST define their AI strategy in line with the IPAO, the data strategy and the three-year plan for information technology in public administrations. This is in order to draw up a comprehensive strategy that is

consistent with the overall policy on the use of technology in public administrations, and which also takes into account the local area in question.

Public administrations MUST implement their AI strategies by entrusting their coordination, together with the data strategy, to the digital transition manager (DTM) and that officer's office (DTO), as provided for in the three-year plan for information technology in public administration 2024–2026, while also involving the Data Protection Officer (DPO) and the Cybersecurity Officer[18] or a similar role.

In contexts where artificial intelligence is used extensively, public administrations MAY establish a dedicated role for AI coordination, the Artificial Intelligence Manager (AIM), who MAY be assigned their own office (AIO), which shall operate in accordance with the DTM's guidelines.

Public administrations MUST monitor and periodically update the strategy in order to:

- verify its effective application and enforceability;
- identify any deviations from the objectives that have been set;
- assess the need for updates in relation to the evolution of objectives and technologies.

4.2. Analysis of the context and characteristics of the public administration

Public administrations MUST analyse the external and internal factors that influence the organisation's ability to achieve the expected outcomes from the use of AI systems, with the aim of identifying an appropriate adoption strategy and model. Public administrations MAY define their strategy and model in collaboration with other administrations or by adopting strategies and models established by higher-level bodies or by administrations of the same type.

External factors include regulatory requirements such as those set out in the DAC, Italian AI Law, AI Act, DGA, GDPR and NIS2, as well as guidelines and decisions adopted by the competent authorities, and developments in technical regulations and technological offerings, with particular reference to the availability

[18] Cybersecurity Officer appointed pursuant to Article 8(2) of Law No 90 of 28 June 2024.

of new AI services and solutions. External factors also include the needs expressed by stakeholders such as the public, businesses and other public administrations.

Public administrations MUST, in relation to processing operations that entail high risks under the AI Act, identify stakeholders and representatives of data subjects in order to assess the potential impacts arising from the use of AI systems, the associated risks to individuals' fundamental rights and freedoms and the appropriate measures to mitigate those risks effectively.

Internal factors include requirements relating to organisational structure (including the size of the body), internal stakeholders (e.g. the body's personnel), administrative procedures, specific operational areas, territorial context and technological capacity. These include the availability of the data, infrastructure and expertise necessary for the effective implementation and management of AI.

In this context, emphasis is placed on the importance of classifying public administration data and digital services, as provided for in ACN Regulation No 21007 on cloud infrastructure and services for public administrations. This classification relates to the minimum standards for the security, reliability, processing capacity and energy efficiency of digital infrastructure for public administrations, as well as to the quality, security, performance, scalability, interoperability and portability of cloud services for public administrations.

Based on analyses carried out at the international level by bodies such as the EU JRC[19] and the OECD[20], operational support has been defined (see the *'Valutazione del livello di maturità nell'adozione di IA'* (Assessment of maturity level in the adoption of AI) tool) to determine the organisational and technological maturity level of PAs.

The actions set out in the AI strategy (see Section 4.1.) are derived from an analysis of the context and characteristics of the public administration, with particular reference to:

- improving the quality of the public administration's assets in terms of data and digital documents;
- strengthening skills;
- establishing collaborations with other PAs for AI trials and management.

[19] EU JRC AI Watch, AI Watch "Road to the Adoption of Artificial Intelligence by the Public Sector", AI Watch "European landscape on the use of artificial intelligence by the public sector. Annex II, Case studies description"
[20] OECD Observatory of Public Sector Innovation

An analysis of the context provides the basis for assessing the feasibility of the objectives and identifying use cases where AI may be deployed effectively, while defining how it should be implemented.

4.3. Objectives and priority areas of application

Public administrations MUST adopt AI technologies by first identifying the objectives and priority areas of application based on their specific context, although they may conduct trials of AI use in non-priority areas that are less prone to risk.

Public administrations MUST carry out a detailed and documented assessment of their needs with the aim of identifying the use cases in which AI offers the greatest benefit in terms of improving operational efficiency and service delivery.

This analysis MUST include an assessment of the economic costs and benefits, taking into account potential savings in resources and improvements in the quality and effectiveness of services.

Assessing specific needs contributes to a potential feasibility study and comparative assessment necessary for selecting the right AI solutions, as provided for in Article 68 of the DAC.

The priority areas for the use of AI by the public administration shall be based on a review and analysis of ongoing projects and trials at national, EU and international level[21].

These areas, however, do not exhaust the range of possible applications: public administrations may also adopt AI solutions in contexts not expressly mentioned, provided that they are deemed relevant to improving services, operational efficiency or decision-making capabilities.

The priority areas identified are as follows.

[21] In September 2024, the Agency for Digital Italy began a review of AI projects within the public administration and of strategic databases for AI purposes, as provided for in the three-year plan for information technology in public administration 2024–2026 (see RA5.4.4 – Development of AI applications of national significance, and RA5.5.1 – Strategic national databases). The information contained in this and the following sections is based on the findings of the aforementioned review and the analyses carried out by the Digital Agenda Observatory at the Politecnico di Milano and by JRC AI Watch on projects involving the use of AI in public administrations in the Italian, EU and international contexts in 2024.

1. **Improving operational efficiency**: public administrations MAY use AI to enhance their data analysis and management capabilities and to automate repetitive processes, with a view to streamlining internal procedures, reducing processing times and improving overall efficiency. In particular, the areas where the greatest benefits are identified are the following.
    a. **Supporting decision-making:** PAs use AI to develop predictive models that enable informed decision-making based on real data, increasing the reliability and timeliness of decisions.
    b. **Optimising resource allocation**: PAs use AI to distribute resources more efficiently, taking into account the social needs of the community, identifying priorities and focusing on the areas of greatest need, thus optimising the use of public resources.
    c. **Improving document management**: PAs use AI to automate the classification, archiving and retrieval of documents, making it easier to search for them and reducing the time spent on administration.
    d. **Improving knowledge management**: PAs use AI to extract relevant information and concepts from unstructured text, thereby improving the accessibility, enrichment and usability of their information assets.
    e. **Improving legal support:** AI can assist PAs with regulatory and case-law analysis, helping them to produce more accurate and timely legal opinions.
    f. **Improving procurement procedures**: PAs adopt AI to optimise procurement procedures at the planning, design, award and performance stages, thereby improving both the efficiency and the transparency of the procurement process.
2. **Improving services for people and businesses:** PAs MAY use AI to enhance their data analysis and management capabilities in order to tailor digital services to users' specific needs, including through a proactive approach. In particular, the areas where the greatest benefits are identified are the following.
    a. **Personalisation:** PAs use AI to tailor public services to the specific needs of people and businesses, improving digital interaction and the efficiency of their responses.

b. **Proactivity:** AI enables PAs to anticipate users' needs, providing relevant services or information before they are requested, thereby simplifying access and reducing waiting times.
c. **Transparency:** PAs use AI to improve transparency, providing people and businesses with clear and immediately accessible information on their obligations and on the progress of administrative procedures initiated with the PA itself.
d. **Accessibility:** PAs adopt AI solutions to make public services accessible and compliant with Article 53 of the DAC and with current legislation and guidelines on the subject, ensuring that digital platforms are usable even by people with disabilities or limited digital skills. In particular, AI must be used as a tool to assist in creating and managing natively accessible content.
e. **Inclusion:** PAs adopt AI to analyse the needs of the public in order to promote services for vulnerable sections of society, promoting social inclusion.

3. **Data security and protection**: PAs MAY use AI to improve data and infrastructure security by identifying potential threats and ensuring advanced protection.

4.4. Technical regulations

When procuring, developing and deploying AI systems, public administrations MUST take into account technical standards at the national, EU and international level.

The AI Act defines the requirements to be met by high-risk AI systems (see Chapter 5.). The European Commission has asked[22] the European standardisation bodies CEN and CENELEC to develop harmonised technical standards[23], in accordance with Regulation (EU) No 1025/2012, which set out concrete approaches to meet these requirements. From the date on which the AI Act comes

[22] On 25 May 2023, the European Commission adopted Decision C (2023)3215 – Standardisation request M/5932 entrusting the European standardisation bodies CEN and CENELEC with the drafting of European technical standards for AI Act compliance.
[23]The technical standardisation work carried out by the CEN/CENELEC Joint Technical Committee (JTC) 21 'Artificial Intelligence' involves the development of technical standards covering requirements for: risk management, data quality and governance, logging and traceability, technical documentation, transparency, human oversight, accuracy, robustness and cybersecurity.

into force (2 August 2026[24]), high-risk AI systems must be made compliant through a quality management system and a conformity assessment prior to being placed on the market, both of which must be based on the aforementioned technical standards.

At the international level, technical standardisation activities relating to AI are coordinated within ISO/IEC by JTC 1/SC 42 'Artificial Intelligence'.

While taking into account the specific requirements of the harmonised technical standards requested by the European Commission for high-risk AI systems, the standards developed within the ISO and ISO/IEC frameworks currently serve as a useful reference for various key aspects of artificial intelligence. However, it will be necessary to consider future adjustments, depending on the publication of the new harmonised European standards by CEN and CENELEC.

In Italy, UNINFO's UNI CT 533 'Artificial Intelligence' is the technical committee of the national standards body, UNI, with responsibility for standardisation in the field of AI, and is a mirror committee of ISO/IEC JTC 1/SC 42 and CEN/CENELEC JTC 21.

Please refer to the *'Norme tecniche in ambito IA'* (Technical standards in AI) tool for a list of the main technical standards on AI and for useful references to consult the progress of technical standardisation in AI at the national, EU and international level.[25]

4.5. Use cases

In light of the points made in the preceding sections, the proposed model for the adoption of AI requires public administrations to identify use cases where AI offers the greatest benefit in terms of improving operational efficiency and service delivery, based on their own AI strategy (see Section 4.1.), an analysis of the context and the characteristics of the organisation (see Section 4.2.), taking into account the objectives and priority areas for the application of AI (see Section 4.3.).

[24] See Article 113 of the AI Act. In any case, the progress of the technical standardisation work in support of the AI Act may be viewed on the website of the CEN/CENELEC JTC 21 'Artificial Intelligence'.

[25] A comprehensive list of technical standards published at the national, EU and international level, along with an update on the progress of ongoing technical standardisation activities, is available on the UNI/CT 533 website.

Public administrations MAY, however, identify specific use cases for trials, pilot schemes and limited-scope, low-risk initiatives by simplifying the model.

Once use cases have been identified, public administrations MUST collect, document and update key information on these use cases and on the AI solutions that implement them, throughout the entire life cycle of the latter, from conception to experimental proof of concept (POC), through to deployment in an operational environment, and finally to decommissioning[26]. The *'Casi d'uso'* (Use cases) tool provides, by way of example, a use case documentation template[27].

Use cases MAY be implemented jointly by several public administrations, including through the trial and development schemes provided for in the three-year plan for information technology in public administration.

In the case of joint development, responsibility for managing the documentation lies with the lead public administration or a designated representative.

Public administrations MUST periodically submit a summary of the documentation describing use cases to AgID, in accordance with the procedures set out in the three-year plan for information technology in public administration.

AgID periodically publishes, in the form of data sheets and in accordance with the scheme defined by the Tools of the AI Guidelines, a selection of use cases derived from the data collected and the communications sent by administrations, with the aim of highlighting best practices and solutions adopted with regard to emerging issues.

4.5.1. AI functionality

Based on the identified use cases, public administrations shall specify the functionalities of the AI systems that implement those use cases.

The review of projects and trials at the national, EU and international level mentioned above in Section 4.3. has also produced a non-exhaustive classification of the functionalities of AI systems. This classification is reported in the *'Funzionalità dell'IA'* (AI functionality) tool.

[26] To determine the technology maturity level of the use case and the solutions that implement it, it is recommended that the technology maturity levels set out in ISO 16290:2013 Space systems – Definition of the Technology Readiness Levels (TRLs) and their criteria of assessment be used

[27] The template for collecting information on use cases is an adapted version of the template used in the ISO/IEC TR 24030:2024 standard.
Artificial Intelligence (AI) - Use cases https://www.iso.org/obp/ui/en/#iso:std:iso-iec:tr:24030:ed-2:v1:en.

4.5.2. Requirements

During the development and management stages of AI systems, public administrations MUST identify requirements based on the regulatory, technical, ethical and operational needs that must be met, taking into account the expectations of stakeholders. Public administrations MUST apply, in addition to the *Principles* set out in these Guidelines (see Section 3.4.), the following categories of requirements:

- compliance with the AI Act (see Chapter 5.)
- data management and quality (see Chapter 9.)
- personal data protection (see Chapter 9.)
- cybersecurity (see Chapter 9.).

When identifying requirements, public administrations MUST also take into account the risk level of the AI system to be implemented (see Section 4.7.). In the case of high-risk systems, the mandatory requirements set out in the AI Act apply.

4.5.3. Key performance indicators

Public administrations MUST define, for each use case, key performance indicators (KPIs) that measure the effectiveness of the AI system in achieving the objectives that have been set.

The KPIs MAY be different depending on the maturity level of the AI system.

The KPIs MUST be monitored throughout the life cycle of the use case, in order to assess the evolution of the performance of the AI system.

KPIs MAY measure both technical performance and the impact on business processes, as well as the added value for the public administration. KPIs may relate to:

- the performance of the model;
- data quality;
- robustness and reliability;
- computational efficiency;
- usability and accessibility;
- ethical impact;
- regulatory compliance;
- cost and sustainability.

The *'Indicatori di prestazione'* (Key performance indicators) tool provides operational support for identifying the KPIs of an AI system.

4.6. Governance

The public administration's governing bodies shall approve the organisation's AI strategy.

In accordance with the provisions of the three-year plan for information technology in public administration, responsibility for the management of AI within public administrations MUST be entrusted to the DTM and their office (DTO), in accordance with the procedures set out in the strategy (see Section 4.1.). The governing bodies MUST ensure that this delegation takes place and MUST communicate it within the public administration.

Specifically, the DTM is granted the responsibility and authority for:

- ensuring that AI systems comply with the provisions of these Guidelines;
- reporting to the governing bodies on the performance, inspection, monitoring and development of AI systems.

Public administrations MUST establish governance procedures for developing and using AI systems. Any procedures MUST be consistent with the organisation's AI strategy and with the strategy set out in the IPAO.

By way of example, and to assist public administrations in their work, examples of possible AI governance procedures are provided in the '*Procedure di governance*' (Governance procedures) tool.

Without prejudice to their obligations under the AI Act, public administrations MAY adopt a code of ethics and conduct for AI. This code SHOULD become a binding governance tool, aligned with the current regulatory framework, integrated into the public administration's decision-making and operational processes and aimed at the responsible, fair and transparent use of AI (see Chapter 6.).

4.7. Risk management

The AI Act sets out strict requirements for the identification, assessment and mitigation of risks associated with 'high-risk' AI systems, in particular to ensure the protection of the health, safety and fundamental rights of individuals.

Public administrations MUST adopt a risk management approach that complies with the AI Act. Pending the finalisation of specific harmonised technical

standards (see Section 4.4.), public administrations MAY refer to the UNI ISO 31000[28] and ISO/IEC 23894[29] technical standards. However, the latter shall be applied taking into account that the ISO standards, if not harmonised, do not guarantee the presumption of conformity provided for in the AI Act.[30]

The '*Valutazione del rischio*' (Risk assessment) tool provides operational guidance to support risk analysis activities for 'high-risk' AI systems.

PAs MUST manage the specific risks arising from the use of AI systems while also taking into account the risks to the fundamental rights and freedoms of data subjects arising from the related processing of personal data, in compliance with the principles of minimisation, integrity and confidentiality, privacy by design and by default and the security obligations provided for by the legislation on personal data protection and the AI Act.

Conversely, for AI systems classified as 'limited risk' or 'minimal or no risk', public administrations SHOULD identify risks and countermeasures associated with the use of such systems. The analysis may be carried out in a simplified form compared to 'high-risk' systems.

It should be considered that AI systems may integrate AI models for general purposes (GPAI). In such cases, following the risk analysis, the system's initial classification may change, making it necessary to reclassify even systems initially assessed as lower risk as 'high risk'. In the event of such reclassification, the specific requirements laid down in the AI Act for high-risk systems shall apply in full.

It should be noted that, subject to regulatory requirements, risk analysis is not an activity recommended solely for the development of AI systems; on the contrary, it is a standard methodological step in most projects, whether IT-related or not. Therefore, the guidance set out in this section should not be interpreted as imposing further obligations that add to the complexity of developing AI systems, which is already a burdensome process in itself. On the contrary, the aim is to apply a process common to all types of project (risk analysis) in such a way that this analysis helps to achieve the intended objectives while minimising undesirable effects.

[28] UNI ISO 31000:2018 Risk Management – Guidelines
[29] ISO/IEC 23894:2023 Information technology -Artificial intelligence -Guidance on risk management
[30] Standards in Europe

4.8. Impact assessment

The AI Act sets out strict requirements for the impact assessment for 'high-risk' AI systems (FRIA, see Chapter 5.) which, where any of the obligations under Article 27 of the AI Act are already met through a data protection impact assessment (DPIA), supplements that assessment.

The results of the impact assessment MUST be documented and shared with stakeholders.

Without prejudice to the obligations under the AI Act, public administrations SHOULD establish a process to assess, even if this is in a simplified form, the potential impact on fundamental rights arising from the development, use or possible misuse of AI systems, including those classified as 'minimal or no risk'. The data protection impact assessment pursuant to Article 35 GDPR and Article 27 of Directive (EU) 2016/680 and the measures of the Italian Data Protection Authority (see Chapter 10).

Pending the establishment at EU level of the tools to support compliance with the requirements set out in Article 27 of the AI Act regarding the FRIA, these Guidelines propose the *'Valutazione Generale di Impatto dell'Adozione dell'Intelligenza Artificiale'* (General impact assessment of the adoption of artificial intelligence) tool or VGIAIA. The VGIAIA serves as a general tool to assist public administrations in carrying out the necessary preliminary AI impact assessment while also incorporating elements relating to fundamental rights. The VGIAIA does not replace the FRIA, which MUST be carried out by the deployer for high-risk AI systems, as provided for in Article 27 of the AI Act.

4.9. Operational plan

Public administrations MUST set clear operational objectives for the adoption and use of AI (see Section 4.3.), ensuring that these are consistent with their AI strategy and comply with regulatory requirements. Objectives MUST be measurable, monitored and updated periodically.

To achieve the objectives, PAs MUST adopt a systematic and documented approach to project management by defining an operational plan, which MUST:

- define the operational activities to be carried out;
- identify the resources required;

- define the timeframe within which the activities must be completed;
- assign specific responsibilities;
- establish criteria for assessing and monitoring the results.

The operational plan MUST include all activities necessary to achieve the objectives relating to the management (see these Guidelines), procurement (see the Guidelines on AI procurement) and development (see the Guidelines on AI development) of AI systems currently in use or to be adopted by the public administration.

4.10. Resources, skills and communication

Public administrations MUST identify and allocate the appropriate financial, technological and human resources to develop, maintain and continuously improve their AI systems.

Resources may include:

- data used in the various stages of the life cycle of the AI system;
- AI algorithms and models, including open-source ones;
- software, including open-source software;
- IT infrastructure (e.g. cloud computing, edge computing and computing resources);
- human resources with the expertise necessary to manage the life cycle of the AI system.

Public administrations MUST determine the skills that are necessary for the personnel carrying out activities relating to AI systems. Public administrations MUST ensure that such persons acquire the necessary skills through training, education or experience, and that they keep records to demonstrate the skills acquired.

Public administrations MUST take the necessary measures to acquire the relevant skills and to assess their effectiveness (see Chapter 8.).

Public administrations MUST promote the responsible and effective use of AI, ensuring that staff are fully aware of the AI principles, objectives and strategy, as well as the code of ethics and conduct.

Public administrations MUST draw up an internal and external communication plan relating to the AI systems they have adopted (see Chapter 7.).

Without prejudice to their obligations under the AI Act, public administrations MUST document and retain the information necessary to ensure the regulatory compliance, transparency and traceability of AI systems. Operational documentation that is useful for ensuring the effectiveness and continuous improvement of the system includes manuals, procedures, impact assessments, monitoring reports and the register of complaints.

Public administrations shall apply the AgID guidelines on the creation, management and retention of electronic documents to such documentation.

4.11. Implementation

At this stage, public administrations MUST implement the operational plan drawn up during the planning phase (see Section 4.9.).

Also at this stage, public administrations MUST carry out an assessment of the risks associated with AI, either at scheduled intervals or in the event of significant changes (see Section 4.7.). Public administrations MUST keep adequate records of the results of assessments to ensure traceability, compliance and timely corrective action.

Public administrations MUST carry out an impact assessment of the AI system in accordance with the established plan (see Section 4.8.) and in the event of significant changes. Public administrations MUST maintain adequate documentation on the results of all assessments carried out.

4.12. Monitoring and assessment

Public administrations MUST define – and then monitor regularly, keeping a record of the monitoring results – useful KPIs:

- to measure the effectiveness of the organisational and technical measures planned by the public administration for AI management;
- to measure the technical performance of AI systems;
- to assess the impact of the system on the PA's internal processes;
- to assess the impact on fundamental rights;
- to assess the added value that the AI system brings to the PA.

4.13. Continuous improvement

Public administrations MUST ensure, over time, that the AI systems they have adopted, as well as the technical and organisational measures for managing AI, remain fit for purpose, appropriate and effective through a process of 'continuous improvement'.

5. Conformity of AI solutions

Public administrations that adopt AI systems MUST act in full compliance with national and EU legislation on AI, with particular reference to the Italian AI Law and the AI Act.

Public administrations MUST refer to the definition of 'AI system' set out in Article 3(1) of the AI Act and referred to in Chapter 3 of these Guidelines in order to determine whether a technology that the public administration has already adopted or intends to adopt falls within the scope of the aforementioned regulation.

Public administrations MUST classify the AI system by determining its risk level, in accordance with the provisions of Article 6 of the AI Act (see Section3.3.).

Public administrations MUST recognise the role they play under the AI Act, for example, provider or deployer (see Chapter3.2.),

In accordance with the provisions of Article 25 of the AI Act, the circumstances in which public administrations act as providers of AI systems include:

- the provision of AI systems through cloud services to other public or private entities;
- the development and making available through reuse of AI systems, including prompt lists.

Determining the risk level to which the AI system belongs and the role played by the public administration in relation to that system is essential for identifying the applicable provisions of the AI Act.

5.1. Responsibilities along the AI value chain

Responsibilities and obligations along the AI value chain must be consistent with the AI Act, which sets out clear obligations depending on the role within the chain, as defined in Articles 22, 23, 24, 25 and 26 of the AI Act. Specifically, as set out in Article 25:

Any distributor, importer, deployer or other third-party shall be considered to be a provider of a high-risk AI system for the purposes of this Regulation and

shall be subject to the obligations of the provider under Article 16, in any of the following circumstances:

- they put their name or trademark on a high-risk AI system already placed on the market or put into service, without prejudice to contractual arrangements stipulating that the obligations are otherwise allocated;
- they make a substantial modification to a high-risk AI system that has already been placed on the market or has already been put into service in such a way that it remains a high-risk AI system pursuant to Article 6;
- they modify the intended purpose of an AI system, including a general-purpose AI system, which has not been classified as high-risk and has already been placed on the market or put into service in such a way that the AI system concerned becomes a high-risk AI system in accordance with Article 6.

5.2. Monitoring the life cycle of AI solutions

Public administrations, whether acting as providers or deployers, MUST monitor their AI systems and the associated impact on fundamental rights throughout the system's life cycle. The type of monitoring (in terms of activities, frequency and effort) depends on the level of risk identified during the classification process and on the role played by the public administration (as a provider or deployer).

Without prejudice to the obligations set out in the AI Act for 'high-risk' systems, public administrations MUST, even for AI systems classified as 'limited risk' or 'minimal risk':

- verify that the data is processed in accordance with the regulations in force, with particular attention to personal data protection, copyright and any other sector-specific regulations;
- ensure the presence and exhaustiveness of the technical documentation supporting the AI system; the documentation MUST include elements that are useful for verifying the principles of transparency and explainability of the outputs; the documentation MUST be updated throughout the life cycle of the AI system.

Monitoring activities may consist, for example, of:

- **collection and verification of documentation:** compilation of reports and studies describing the characteristics of AI systems, their impact on public administration services and on fundamental rights;
- **operational checks and tests:** testing of AI systems in real-world conditions to assess their performance, reliability and any potential issues;
- **collection and analysis of complaints:** structured compilation of reports from users, members of the public or the PA's staff, including a qualitative and quantitative analysis of the issues highlighted;
- **exchange of information and best practices:** creation of a periodic comparison system with the involvement of other PAs, providers, universities and research centres to monitor and improve the use experience.

For high-risk AI systems, public administrations, in their capacity as deployers, MUST carry out a fundamental rights impact assessment (FRIA) in the cases specified in Article 27 of the AI Act, prior to the system's deployment. When carrying out this assessment, public administrations MUST take the following factors into account:

- the **description of the PA's processes** in which the high-risk AI system will be used;
- the **period and frequency** with which the high-risk AI system will be used;
- the **categories of natural person and groups** likely to be affected by its use in the specific context;
- the **data being processed**, paying particular attention to the presence of personal data and, where such data is present, carrying out a specific assessment;
- the **specific risks of harm** that may affect the categories of natural persons or groups of persons affected by its use and the related fundamental rights;
- the **measures to be taken** should such risks materialise, including provisions relating to internal governance and complaints mechanisms;
- the description of **the human oversight measures**.

If, on the other hand, a public administration acts as a provider of high-risk AI systems, it MUST do the following.

- **Document the post-market monitoring system**, which must be proportionate to the nature of the AI technologies and the risks associated with the high-risk AI system. The post-market monitoring system collects, documents and analyses relevant data gathered throughout the product's entire life cycle. Post-market monitoring may include analysis of interaction with other AI systems.
- **Monitor** the definition of and compliance with the risk management system. This process aims to identify and mitigate the relevant risks posed by such AI systems to health, safety and fundamental rights. The risk management system MUST be periodically reviewed and updated to ensure its continued effectiveness, as well as the justification and documentation of any significant decisions and actions taken.
- **Make timely and comprehensible information available** in relation to how these systems are developed and operated.
- Ensure, in line with the provisions of Article 12 of the AI Act, at a technical level, that the **recording of relevant events** (logging) takes place automatically, for the entire life cycle of the system.
- **Keep records** and have available the technical documentation containing the information necessary to assess the system's compliance with the requirements for high-risk AI systems.

5.3. Oversight measures

Public administrations MUST comply with the obligations regarding human oversight measures for high-risk AI systems set out in Article 14 of the AI Act, in accordance with their respective roles.

- The PA provider MUST design and develop AI systems in a manner that ensures effective human oversight, incorporating appropriate interfaces and control tools.
- Public administration deployers MUST adopt procedures and tools to ensure the monitoring, interpretation and, where necessary, human intervention in decisions generated by AI.

By way of example only, and in accordance with the standards and best practices developed by the relevant international and EU organisations, public administrations SHOULD adopt human oversight measures that allow, where possible, for intervention in the parameters that determine the AI system's output ('human-on-the-loop'), for influencing that output ('human-in-the-loop') or to determine how it is used in the public administration's decision-making processes ('human-in-command').

Public administration deployers shall entrust the human oversight of high-risk AI systems to the DTM and their office (DTO) and, as already clarified above, MUST ensure that the staff assigned to this task possess the necessary skills, experience and authority to perform this function effectively.

5.4. Record keeping

Public administrations MUST ensure that documentation relating to the use of high-risk AI systems is correctly retained, in accordance with Italian and EU legislation on document management. This includes the storage of the data and metadata necessary to ensure the traceability of automated decision-making processes, in compliance with the legislation on personal data protection.

The AI Act (Article 18) establishes an obligation to retain technical documentation on high-risk AI systems, including a description of their operation, performance and the monitoring and control procedures put in place to ensure compliance with the legal requirements. Such documentation MUST be retained for a period appropriate to the use of the system, so as to ensure that audits and inspections may be carried out by the competent authorities, taking into account the retention period for documentation relating to the process supported by the AI system.

Public administrations MUST manage and retain documentation relating to AI systems in accordance with the provisions of the DAC and in compliance with the AgID Guidelines on the creation, management and retention of electronic documents.

6. Ethical governance of AI

Without prejudice to the obligations to comply with the AI Act (see Chapter 5.) and any other applicable EU and national legislation, including sector-specific regulations, which MUST be strictly adhered to, public administrations MAY adopt codes of ethics and conduct. Such codes may also be developed in collaboration with other public administrations or by adopting models established by higher-level bodies or bodies of the same type.

The content of codes of ethics and conduct may include provisions setting out additional elements and requirements for developing and using AI systems, in addition to those already required under the applicable legislation. In particular, pursuant to Article 95(1) of the AI Act, codes of ethics and conduct should encourage the voluntary application of some or all of the requirements set out in Chapter III, Section 2 of the AI Act to all AI systems other than high-risk AI systems adopted by the public administration.

Codes of ethics and conduct, as tools designed to promote a shared culture of responsibility in the use of AI, tend to:

- assist PAs in selecting and using AI systems by incorporating and implementing the relevant regulations applicable to the development or use of AI systems;
- introduce guidelines on the use of AI systems by employees; these guidelines must be shared with management and/or senior executives.

Codes of ethics and conduct CANNOT, under any circumstances, result in a reduction in safeguards or the removal of obligations laid down by current legislation.

Public administrations MAY develop codes of ethics and conduct, in accordance with Article 95(3) of the AI Act, in their capacity as providers or deployers, and should, in the process of drafting such codes, encourage the involvement of stakeholders, including private-sector stakeholders (organisations representing civil society, universities, research centres, provider associations and regulatory bodies).

When identifying any additional elements and requirements, public administrations MUST take into account:

- the available technical solutions and the sector's best practices;

- the ethical principles that have been developed and/or are currently being developed in this area, as set out in Article 95 of the AI Act.

To this end, public administrations MUST take due account of the **Ethics guidelines for trustworthy AI** issued by the European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG)[31].

Public administrations MUST also take into account the main EU or international sources or initiatives when developing and consolidating their codes of ethics. Examples of relevant initiatives, though this list is by no means exhaustive, include documents drawn up by:

- The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, a group promoting initiatives in the field, for example, *Ethically Aligned Design*[32]*;*
- The IEEE Standards Association, a body that contributes to the development of global standards promoting responsible innovation in AI and autonomous systems; notable standards include the P7000 *Series*[33] and the P7010 *Series*[34];
- The G7 Hiroshima AI Process, launched by the G7 under the chairmanship of Japan, is a reference framework aimed at promoting safe, secure and reliable artificial intelligence[35];
- The Organisation for Economic Co-operation and Development, which has developed the AI Principles recommendations on AI.[36]
- The United Nations Educational, Scientific and Cultural Organization (UNESCO), which has developed the [37]*Recommendation on the Ethics of Artificial Intelligence.*

[31] Ethics guidelines for trustworthy AI https://data.europa.eu/doi/10.2759/640340.
[32] l'Ethically Aligned Design https://sagroups.ieee.org/global-initiative/wp-content/uploads/sites/542/2023/01/ead1e.pdf.
[33] IEEE, Standard Model Process for Addressing Ethical Concerns during System Design https://standards.ieee.org/ieee/7000/6781/.
[34] IEEE, Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being https://standards.ieee.org/ieee/7010/7718/.
[35] G7 Hiroshima Process International Code of Conduct for Advanced AI Systems (2023) https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems.
[36] OECD, Recommendation of the Council on Artificial Intelligence (2024) https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.
[37] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2024) https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

- The United Nations Secretary-General's High-level Advisory Body on Artificial Intelligence, which has drawn up the report entitled *Governing AI for humanity.*[38]
- Vatican City State, *Rome Call for AI Ethics*[39].
- Pontifical Commission for the Vatican City State, Decree No DCCII, *Guidelines on Artificial Intelligence*[40].

When drafting a code of ethics, public administrations MUST pursue a series of objectives (in addition to those already explicitly set out and protected by the*AI Act*). In particular, such codes should 'guide the conduct of those involved in the development and adoption of AI systems, setting out guiding principles and values, with a view to promoting social and environmental well-being through AI', emphasising ethical objectives such as:

- ensuring high *standards* of professional skills and ethical practice;
- promoting the continuous training of staff involved in AI in order to develop the technical skills and capabilities needed to address the ethical challenges related to the use of such technologies;
- promoting public awareness, among citizens, of AI technologies and their consequences;
- using AI to tackle global challenges and safeguard the public good;
- investing in research to mitigate security and societal risks;
- contributing to the development and adoption of international technical standards for AI.

Public administrations MUST ensure that codes of ethics are reviewed periodically in order to ensure they are kept up to date with technological and regulatory developments and to align these initiatives with the *SDGs* in line with the *European Green Deal*. Reviews MUST be carried out in consultation with experts, public- and private-sector representatives and citizens' rights associations.

The '*Modello di codice etico*' (Model code of ethics) tool provides operational support which, in the context of the general principles relating to AI ethics, can

[38] ONU, Governing AI for Humanity (2024) https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.
[39] Vatican City State, Rome Call for AI ethics https://www.romecall.org.
[40] Vatican City State, No DCCII – Decree of the Pontifical Commission for the Vatican City State, 'Guidelines on Artificial Intelligence' https://www.vaticanstate.va/images/N.%20DCCII.pdf

help guide and support PAs in drawing up codes of conduct with content that is consistent with the relevant regulations.

7. Communications

7.1. Transparency measures

Public administrations that adopt AI systems MUST implement transparency measures designed to ensure that users are aware of their use and of the impact such systems may have on decision-making processes. Transparency is a fundamental requirement for ensuring trust and confidence in the technologies used, in accordance with the principles of transparency and administrative accountability laid down in EU regulations.

The AI Act stipulates that high-risk systems and certain AI systems (AI Act, Article 50) are subject to stringent transparency requirements. Public administrations, in their capacity as providers or deployers, MUST apply these requirements or ensure that they are applied. These include the following:

- **Information on interaction with AI:** providers MUST design and develop AI systems in such a way that users are informed when they are interacting with an AI system. This means that people must be aware of when they are communicating with a machine instead of a human being.
- **Indication of artificially generated content:** AI systems that generate synthetic content, such as text, images, audio or video, MUST clearly identify such content as artificially generated.
- **Technical documentation:** providers of AI systems, in particular those considered high-risk (e.g. biometric recognition, assessment systems for access to essential services) MUST maintain technical documentation describing the functioning of the system; this documentation must be made available to the competent authorities in case of need.
- **Requirements for high-risk systems:** providers MUST ensure that AI systems are designed to provide adequate transparency, including detailed instructions for those responsible for their use, and that high-risk systems allow for the traceability of the decisions made, so that users understand how the system processed the data and reached a particular decision.

The main objective of these requirements is to ensure that users are always aware of the possible use of automated systems in their interactions with PAs and

that they may exercise their rights (e.g. access, rectification and objection), in compliance with the principles of legality and transparency.

Public administrations MUST be transparent in their use of AI systems that process personal data, providing adequate, comprehensive and accessible information to data subjects regarding how these systems influence decisions and what the consequences are for users of public services, and making public the purposes, criteria and methods used by the AI systems. In addition, they must ensure that the information provided is accessible to all interested parties, including through the use of plain and simple language.

7.2. Privacy notice requirements

In accordance with the GDPR and the AI Act, and as mentioned above, public administrations MUST ensure a high level of transparency and clarity in the use of AI systems, particularly where such systems may have an impact on users' rights and interests.

In this context, it is essential to provide clear, comprehensive and accessible information so that users understand how the AI systems used by the public administration work and the rights that this entails for data subjects. In addition to the requirements set out in Articles 12, 13 and 14 of the GDPR regarding the processing of personal data, the privacy notice must include the following.

- **Description of the AI system and its purposes:** public administrations MUST provide a detailed description of the AI system used, specifying its nature and objectives. The notice must clarify:
  - **type of AI used:** whether it is a machine-learning system, a predictive analytics algorithm or another model based on AI;
  - **purpose of the system:** PAs MUST clearly indicate the purposes for which the system is used, specifying whether the AI solution is used to improve the services offered, to automate administrative processes or for other purposes related to improving administrative efficiency and effectiveness.
- **Decision-making criteria and impact on data subjects:** public administrations MUST clearly explain the criteria and parameters on which

the automated decisions made by the AI system are based. It is essential that the information includes:

- **how the system works,** specifying whether the AI solution operates by analysing historical data or whether it uses statistical models to make predictions or recommendations;
- **possible impacts of automated decisions,** describing the potential impact of decisions made by the AI solution on the public, in terms of both their rights and any practical consequences, such as their ability to access certain services or benefits.

- **Explanation of automated decisions:** citizens MUST be informed of their right to receive a comprehensible explanation regarding decisions made by AI systems; public administrations MUST therefore:
  - **ensure that the decision-making process is transparent:** provide, upon request, clear and simple information enabling citizens to understand the rationale behind automated decisions affecting them;
  - **describe the criteria and data used:** be able to explain what data was used in the process and how that data influenced the outcome of the decision.
- **Right to object and human intervention in the decision-making process:** it is essential that public administrations inform their users of their right to object to automated decisions and to request human intervention in the decision-making process, where such decisions could have a significant impact on their rights. For this purpose, it is necessary to:
  - **specify the procedures for exercising the right to object:** public administrations MUST inform citizens of the procedures and channels through which they may object to an automated decision, specifying the response times and the procedures for requesting a review by a human operator;
  - **ensure access to human intervention:** if a citizen believes that an automated decision is inaccurate or unfair, they MUST be given the opportunity to request human intervention to review the decision and, if necessary, amend it.

Public administrations SHOULD draw up detailed, tailored information sheets for each AI system used, adapting the communication to the specific context and

user needs, and prioritising the use of clear, simple language free from technical jargon.

7.3. Adopting AI in institutional communications

Communications are a strategic tool for public administrations, enabling them to keep people and businesses informed and to promote public services. Thanks to advanced technologies such as natural language processing, machine learning and automation, AI can enhance the effectiveness, accessibility and interactivity of communications, contributing to a more transparent and inclusive public service.

AI can therefore support public communications, although this requires careful management of the ethical and operational challenges involved.

The opportunities and advantages that AI brings include:

- optimisation of creative processes: analysing trends and data in real time, generating textual and visual content and creating targeted and innovative communication campaigns;
- personalisation: personalising user experiences and making each interaction unique and meaningful;
- automation and rapid responses: chatbots and virtual assistants respond immediately and reduce waiting times;
- improved accessibility: information can be made more accessible to citizens with disabilities through the use of translation tools or text-to-speech software;
- cost reduction: by automating repetitive tasks, valuable time is freed up for creative strategy.

Public administrations that adopt AI systems for institutional communications MUST follow a gradual and planned approach, under the coordination of the digital transition manager (DTM), in accordance with the provisions of Chapter 4.. Furthermore, as already mentioned in Chapter 8., it is important that employees receive appropriate training. PAs adopting AI in their communication activities:

- MUST include a section in the annual communications plan dedicated to the use of AI in communications activities;
- MUST monitor and assess the effectiveness of the AI tools used;
- MUST train staff on the use of AI tools;

- MAY adopt AI pilot schemes to assess their impacts and actual usefulness.

Public administrations may use AI-based tools and technologies to enhance their communications activities. In particular, though this list is not exhaustive, public administrations:

- MAY use chatbots and virtual assistants that improve interaction with users and reduce response times;
- MAY use sentiment analysis tools for social media and other forms of interaction;
- MAY adopt social media monitoring and management platforms to automatically manage the publication of content and personalise content;
- MAY use automatic natural language processing (NLP) tools to summarise complex documents or answers to questions;
- MAY use machine translation systems to make content accessible to a wider audience.

The use of AI in communications can bring numerous benefits, but it is essential to adopt a responsible and transparent approach. In particular, PAs using AI for their communications activities:

- MUST inform citizens about the use of AI and how their data is managed;
- MUST ensure the protection of any personal data that may be processed;
- MUST be transparent and indicate the use of AI tools;
- MUST ensure maximum accessibility and comprehensibility for everyone;
- MUST NOT develop excessive dependence on AI or fully rely on AI, always ensuring human oversight;
- MUST NOT adopt impersonal communications, but maintain a human and empathetic tone.

8. Training and skills development

In order to fully realise the benefits of AI, public administrations MUST put in place a number of enabling factors – managerial, organisational and cultural as well as technological – among which the acquisition and development of appropriate skills at both the individual and the organisational levels plays a key role.

Public administrations MUST be able to fully understand the potential and implications of AI in order to improve the quality, accessibility and efficiency of public services, for example through the automation of routine tasks, predictive analytics and decision-making support.

Furthermore, public administrations MUST develop specific skills to govern and regulate the use of AI, ensuring compliance with ethical principles, personal data protection and transparency. This is essential for promoting the responsible and sustainable adoption of these technologies, which can help to improve equity and inclusion in the delivery of public services.

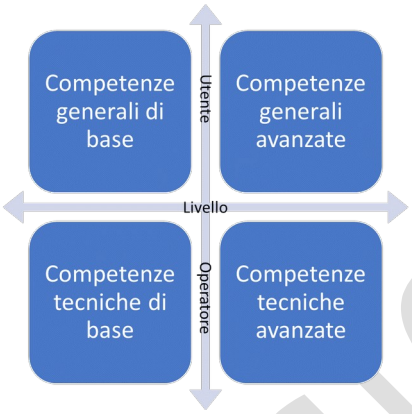
Finally, the ability to provide training and skills development opportunities in AI can position the public sector at the forefront of technological innovation and make it more attractive to talent, helping to retain the best staff and strengthen the competitiveness of the administrative system as a whole.

Article 4 of the AI Act, entitled 'AI literacy', highlights the need for tailored training programmes based on the background, the role undertaken and the intended use of AI solutions. This is to ensure that all individuals (or stakeholders) who interact in various ways with AI-powered solutions can do so in an informed, safe, reliable and effective manner. As also highlighted in the OECD's AI Toolkit[41], just like any other digital transformation within the public sector, achieving maturity in the use of AI requires a broad and diverse range of skills at various levels of the organisation, spanning from core to specialist skills, and calls for the development of tailored and targeted upskilling and training initiatives.

In general, it is therefore necessary to carefully and comprehensively map the professionals present in PAs, with the aim of clearly identifying the level of training needs and the responsibilities related to AI governance. This process will enable targeted training programmes to be developed, ensuring that staff have the necessary skills to manage AI systems effectively and responsibly. The mapping

[41] OECD/UNESCO (2024), G7 Toolkit for Artificial Intelligence in the Public Sector, OECD Publishing, Paris, https://doi.org/10.1787/421c1244-en.

should at least provide evidence of an initial level of needs, which includes core digital skills, including AI literacy, that is distinct from a more advanced level of needs. In addition to this classification, a distinction must be made between general competencies, which encompass the essential characteristics of AI applications, and technical and domain-specific competencies.



Competenze generali di base	Core general skills
Competenze generali avanzate	Advanced general skills
Competenze tecniche di base	Core technical skills
Competenze tecniche di avanzate	Advanced technical skills
Utente	User
Livello	Level
Operatore	Operator

As regards core general skills, these are the skills needed to ensure the ability to interact with AI applications, which are becoming increasingly widespread in many aspects of social and working life (including the public sector). This pervasiveness highlights the need to broaden the range of core digital skills among all public-sector employees, including a basic understanding of the existence of AI and its potential to influence their work (AI awareness). Civil servants need to recognise the key features of AI applications developed on a bespoke basis or integrated into the latest versions of the management and office automation systems provided by their organisation, in order to understand the opportunities they offer and how to use them, and to make the most of their potential to make their work faster, more efficient, more secure and more informed. At the same time, they need to understand the implications of applying the ways in which they interact with AI applications in their private lives to the workplace (e.g. AI-powered search engines, generative AI applications based on individual licences, embedded applications, on social media, etc.). Full awareness of the optimal methods and conditions for using AI applications, in addition to

ensuring the correct adoption of such systems, is also extremely important for overcoming the physiological resistance that the PA's digital transformation path has notoriously encountered. In the case of AI, such resistance may be further reinforced by an incorrect interpretation of the regulatory focus on the risk-based approach promoted by the AI Act. With this in mind, in order to make the most of the opportunities on offer, all public sector employees SHOULD build basic AI literacy, as it has become an essential part of what are known as core digital skills.

At a more advanced level of general skills, the AI literacy framework combines an understanding of the characteristics of AI systems with a thorough grasp of the regulatory, operational and domain-specific aspects associated with AI. Civil servants also need to be trained in the ethical principles relating to the use of AI, including transparency in automated decision-making, and must also develop the skills needed to analyse and manage the security risks to which AI-supported activities and the data used by such systems are exposed.

With regard to technical skills, meaning those required for activities that influence AI performance within an organisation, the ISO/IEC 42001 standard on AI management systems sets out specific requirements regarding the skills of staff involved in such activities, stipulating the need to ensure the ongoing availability of appropriate skills by adopting measures designed to address any identified gaps, and by tracking and monitoring them over time. For obvious reasons of timing, most civil servants and technical managers have rarely had the opportunity to study AI topics in depth as part of their university education, as these fields have developed most significantly in more recent years. It is therefore necessary, among other things, to develop skills relating to modern AI and machine-learning methodologies and technologies, including on the technical aspects of risk-related issues, as well as strengthening the capacity to intercept and manage any crises due to deviations in operating conditions in the event of unwanted events, in a secure and resilient manner.

It should be noted that the strategic and regulatory framework that has developed over recent years also pays particular attention to the issue of upskilling linked to the dissemination and adoption of AI. Public administrations, and in particular digital transition offices (DTOs), need to strengthen and/or develop specialist skills in the field of AI through targeted recruitment and upskilling policies aimed primarily at attracting and retaining talent. This need in the public sector is only partially mitigated by outsourcing design and development activities

through the use of various types of technology partner (in-house bodies or market operators), as it remains essential in any case to possess the appropriate skills to define the organisation's requirements and manage partnerships, all the more so with the consolidation of agile development methodologies, which technology providers are increasingly adopting. It is therefore advisable to identify and strengthen technical and domain-specific skills relating to specific processes involving AI (e.g. training itself or recruitment) and/or areas of application for AI, such as healthcare, transport or education.

The field of AI requires knowledge spanning various disciplines, such as computer science, engineering, mathematics and technology law, as well as sociology, anthropology and philosophy, with particular reference to logic and ethics. The ISO/IEC 42001 standard recommends considering the need for diversified skills according to the different stages of the AI life cycle, as well as ensuring adequate representativeness with reference to the data used for training machine-learning models.

To this end, the IPAO and the staffing requirements plan MUST take into account the need for specialist AI staff.

The professional profiles involved in developing and managing AI-based projects and solutions are very varied. Increasingly critical professional figures include the following.

- ***Project managers:*** professionals who are essential for coordinating and leading projects which include the introduction of AI systems.
- ***Data engineers:*** experts in data design, development and management who will ensure that data is collected, refined and made available efficiently and reliably. Data quality is crucial for returning reliable outputs.
- ***Machine-learning engineer:*** an expert in machine-learning algorithms and programming who designs, implements and optimises machine-learning algorithms that are essential for the digital services developed using this technology to function correctly.
- ***Prompt engineer:*** the expert who trains generative AI to produce effective outputs in accordance with the instructions provided.
- ***Deep-learning engineer:*** an expert in neural networks, deep learning and programming who applies artificial intelligence to more complex problems that require significant computational power through machine

learning, such as speech recognition, computer vision and natural language processing.

- ***Data scientist:*** an expert in data analysis, machine learning and statistics who, working closely with machine-learning engineers, analyses large volumes of data, creates predictive models and works to extract insights using AI and machine-learning algorithms. The data scientist transforms data into value, allowing informed decisions to be made based on concrete, reliable data.
- ***Data steward:*** an operational role responsible for implementing data management policies. In particular, the data steward must: ensure that the data complies with the characteristics defined by the data owner; guarantee the quality and integrity of the data; ensure that the data is used in accordance with the organisation's policy and relevant rules (e.g. regarding data security and protection), including any requirements arising from legislation.
- ***AI engineer:*** an expert specialising in the creation and implementation of artificial intelligence models through theoretical research in the field of AI. This professional figure is crucial for choosing and assessing the right AI model to be applied when implementing software services. Each model has its own characteristics, making it suitable for specific scenarios, so choosing the right one is crucial for the success of the service.
- ***AI architect:*** an expert who assesses and designs the architecture of artificial intelligence systems, ensuring that they are scalable, secure and high performance. Where necessary, they are involved in integrating AI into existing infrastructure. The choice of architecture influences the implementation of the software solution, affecting timeframes, costs and performance.
- **Cybersecurity expert:** an IT security expert who not only analyses and prevents possible cyber threats but also combines the topic of cybersecurity with artificial intelligence, ensuring increasingly efficient defences that are able to autonomously adapt to external threats.
- ***AI ethicist:*** an expert tasked with conducting ethical assessments of artificial intelligence technologies, applications and practices, identifying, analysing and making decisions on value priorities, defining guidelines and providing recommendations to operationalise moral values so as to

translate them into legal standards, particularly on issues relating to discrimination, accountability, transparency and privacy. A professional with expertise in conceptual analysis and in moral norms, principles and reasoning, who facilitates discussions and encourages debate to improve decision-making capabilities in the field of artificial intelligence, while raising awareness among those who develop and use artificial intelligence tools, as the latter must ultimately take full responsibility for them. This person should also keep a close eye on developments in AI research and work to disseminate this knowledge in order to establish a shared 'ethical vocabulary'.

- **IT lawyer:** an expert who combines legal and IT expertise to provide legal advice on the use of IT technologies that may impact on topics such as personal data protection, IT security and intellectual property. A professional who also drafts and reviews contracts relating to software, licences, etc., ensuring that they comply with current regulations, while also addressing issues relating to compliance, dispute resolution and the analysis and assessment of risks arising from the use of technologies such as artificial intelligence.
- **Personal data protection expert:** an expert who ensures that processes, systems and applications using artificial intelligence respect the fundamental right to the protection of personal data and comply with national and international personal data protection regulations, such as the GDPR. A professional who minimises the risks associated with the collection, processing and storage of personal data within AI systems.
- ***Change manager:*** a professional who assesses the impact of AI implementation on sociomaterial structures and develops the necessary change strategies.
- ***Organisational designer:*** an expert in organisational dynamics who provides guidelines for the productive interaction between human and artificial intelligences.

There is therefore a need within public administrations to define the professional profiles and career paths of AI managers and specialists in order to provide appropriate incentives and guidance for recruitment and upskilling initiatives. This is the focus of the work being carried out by UNINFO's technical

committee, UNI/CT 526 'Unregulated Professional Activities', to define professional profiles in the field of AI.

Alongside digital transition managers and their teams, public sector managers also require the appropriate skills and knowledge to make informed, reliable decisions regarding which and how many AI-powered tools to adopt, given their pivotal role in decision-making processes relating to the design and delivery of public services, in compliance with and pursuant to standards and regulations, and more broadly in the process of transforming the public sector.

In particular, public sector managers need to understand the project proposals put forward by experts and be able to assess their potential, implications and impacts[42], combining the core skills shared by all non-IT public sector employees with a full understanding of principles, concepts and applications from not only a technical but also – if not above all – a strategic perspective, as well as the accessibility, legal and ethical implications, in order to fully identify the opportunities and challenges associated with AI. Furthermore, in line with the provisions of the Guidelines for the safe development of AI[43], specialists responsible for implementing AI must develop the ability to understand the threats, risks and associated mitigation measures. Finally, in order to properly guide innovation, decision-makers need to have a full understanding of the complexity factors associated with the implementation of AI solutions and, as with any other change project, the transferable soft skills, managerial skills and leadership skills needed for the process transitioning administrations – first and foremost the digital transition.[44]

8.1. Operational guidelines for PAs

The development of training programmes for all staff, ongoing specialist training for the DTO and the identification and selection of the skills required to manage operations is a strategic process that demands full awareness on the part of the public administration and the appropriate planning of initiatives.

[42] ARISA (2023). AI Skills Needs Analysis, https://aiskills.eu/wp-content/uploads/2023/06/ARISA_AISkills-Needs-Analysis_V1.pdf

[43] The guidelines are published on the ACN website at https://www.acn.gov.it/portale/linee-guida-ia

[44] See, in this regard, the Decree of the Ministry of Public Administration of 28 September 2022 adopting the Guidelines on access to management positions in the public sector, and the Decree of the Minister for Public Administration of 28 June 2023 adopting the Framework of transferable skills for non-managerial staff in public administrations, as well as the Directive of the Minister for Public Administration of 28 November 2023, which introduced a specific framework of leadership skills that managers are required to adopt and against which they must be assessed.

It is a process that involves assessing training needs, setting objectives, designing training programmes, engaging and motivating stakeholders, and carrying out continuous assessment and updating, within planning frameworks (see IPAO) that are well known to public administrations.

In order to build up the necessary skills to implement AI-based applications, public administrations are required to introduce a series of measures in line with their own strategy, including by taking advantage of the opportunities offered by system-wide initiatives that have been implemented or are planned in this area, primarily driven by the National Recovery and Resilience Plan (NRRP).

In particular, based on the analysis and description of the comprehensive system of AI-related responsibilities that spans all levels of the organisation, a set of guidelines shall be drawn up to guide and direct the actions of individual administrations.

Public administrations are also committed to providing training programmes for staff whose roles are partially or entirely replaced by artificial intelligence, with a view to developing skills that will enable them to be redeployed in other areas of the public administration.

Public administrations MUST promote AI literacy among all public sector employees, which is recognised under the MiPA Training Directive 2025 as a core digital skill in every respect[45], by setting and monitoring specific individual performance objectives linked to the development of these skills. In particular, public administrations must encourage staff to achieve the training objectives set out in the Training Directive issued by the Minister for the Public Administration on 24 March 2023[46] for 2024 and 2025, with reference to the training programme available on Syllabus entitled 'Competenze digitali per la PA' (digital skills for public administration) and to use, again on Syllabus, training programmes specifically concerning the application of artificial intelligence by managers and non-IT-specialist staff.

Public administrations MUST ensure that the AI Guidelines produced by AgID are widely disseminated among decision-makers who are involved in, or potentially affected by, the implementation and management of projects and solutions involving the use of AI, depending on their specific skills, so that informed

[45] Directive of the Minister for Public Administration, 'Empowering people and creating public value through training. Principles, objectives and instruments' of 14 January 2025.
[46] Directive of the Minister for Public Administration, 'Planning of training and development of skills relevant to the digital, ecological and administrative transition promoted by the National Recovery and Resilience Plan' of 23 March 2023.

and consistent decisions can be made that address all the technological, managerial, organisational, ethical and legal aspects involved.

Public administrations MUST promote and encourage the participation of DTMs, DTOs and other key stakeholders in the implementation of digital transformation in learning communities, including the communities of practice provided for in the three-year plan for information technology in public administration, on the adoption of artificial intelligence, in order to foster capacity building and the sharing of best practices among administrations. The opportunity for pioneering public administrations to share their experiences, knowledge, methodologies and case studies could prove an effective way of strengthening the public sector's knowledge base and fostering a learning environment that encourages the adoption of AI and willingness to replicate or scale up successful projects. In particular, public administrations MUST promote and encourage the participation of the DTO in the community established by AgID and provided for in the three-year plan for information technology in public administration, which is the responsibility of DTMs (the ReTe Digitale project) and serves as the primary forum for sharing, collaboration and discussion on AI. Public administrations may also encourage the establishment of / participation in other learning communities dedicated to exchanging, discussing and sharing information on specific topics related to these technologies or specific sectors, including those limited to particular types of organisations.

Public administrations that adopt or intend to adopt AI projects and solutions MUST promote and encourage the setting of performance objectives for managers linked to their participation in training programmes designed to support decision-making and to guide AI adoption processes, combining AI literacy with an in-depth exploration of technical, regulatory or ethical issues, and the development of managerial, leadership and soft skills to support change management, including through the use of innovative teaching methodologies. To this end, public administrations may use not only the training programmes available on the Syllabus platform, but also the training programmes offered by the National School of Administration, its regional training centres and Formez PA, as well as those offered by universities, including through the 'PA 110 e lode' programme funded by the Department of Public Administration of the Prime Minister's Office, as well as other training initiatives funded by EU, national or regional funds and self-financing.

PAs implementing AI-based digital transformation projects MUST promote the implementation of targeted and contextualised training programmes to support adoption and change management processes. While AI literacy skills are essential for complementing the key digital skills required of public sector employees and for fostering a common, shared culture that enables digital transformation, the effective roll-out of targeted AI implementation projects within individual public administrations or specific types of administration (such as projects involving healthcare organisations) requires further investment in training and change management to explore specific aspects of design, operation and adoption in more depth. For this purpose, public administrations can develop specific training programmes, including innovative ones (based, among other things, on seminars and conferences, hackathons and mentoring schemes), drawing on projects promoted under the NRRP, such as the PerformaPA initiative, as well as EU, national or regional funds and self-financing.

Public administrations involved in the development, procurement and management of AI MUST facilitate access to specialist technical training programmes aimed at DTOs and IT professionals in general, based on a systematic and regular assessment of skills requirements, including through collaboration with universities (e.g. the PA 110 e Lode project), higher education institutions and centres of expertise at national and international level, and by promoting the acquisition and maintenance of professional certifications.

Public administrations MUST also establish procedures relating to the cybersecurity of AI and concerning the ability to respond to cyber threats, incidents and crises in the context of AI.

Public administrations MUST provide appropriate evidence of each of the training initiatives identified in the training planning section of the IPAO, in accordance with the procedures and deadlines set out for planning, monitoring and reporting in the MiPA Directive on Training 2025.

Public administrations SHOULD promote and encourage a culture of continuous learning focused on training and the ongoing development of artificial intelligence skills. This approach fosters an environment in which professional development is valued and encouraged, making learning an integral part of everyday working life.

Public administrations that use or intend to use AI-based solutions MAY take steps to attract talent and young specialists by promoting initiatives that include,

among other things, hackathons, internships, apprenticeships and PhD programmes. In this regard, public administrations may also take advantage of the funding opportunities offered by initiatives such as the innovative PhD programmes promoted by the Ministry of Universities and Research or the InPA internships and PhD programmes promoted by the Department of Public Administration, or promote collaborative initiatives with universities and higher technical institutes (ITS).

Public administrations that use or intend to use AI-based solutions MAY promote interdisciplinary research through funding initiatives or by participating in dedicated projects in collaboration with training institutions, universities, research centres and AI solutions providers.

Finally, since the adoption of AI cannot ignore the development of digital and artificial intelligence skills of the recipients of the public services that use it, administrations, and in particular local authorities and service-providing administrations, MAY contribute to promoting AI awareness and AI literacy among citizens so that they can interact effectively, consciously and safely with AI solutions. Citizens must be empowered to understand and make the most of the potential and implications of these technologies so that they can increase their use of these services in an informed, safe manner that respects their rights. To this end, public administrations, and in particular local authorities, may promote, among the relevant members of the public (e.g. through information campaigns and by establishing partnerships with implementing administrations), the initiatives funded by the NRRP relating to the Digital Civil Service and the Network of Facilitation Services, in order to raise awareness of the issues surrounding artificial intelligence and how to access services that utilise such technologies[47], as well as launching ad hoc projects.

[47] The Network of Digital Facilitation Services project is coordinated by the Department for Digital Transformation and implemented in collaboration with the regions, autonomous provinces and local authorities. The Digital Civil Service project is run by the Department for Digital Transformation in collaboration with the Department for Youth Policy. Both projects, which operate within Measure 1.7 of the NRRP, albeit with differences in governance and operation, aim to activate centres distributed throughout the national territory where citizens can access facilitation and training services to develop and strengthen their core digital skills. https://repubblicadigitale.gov.it/portale/progetti-del-dipartimento

9. Data management and quality

Recital (67) of the AI Act highlights the importance of data quality as an essential prerequisite for reliable AI systems[48]. The text goes on to note that '*High-quality data sets for training, validation and testing require the implementation of appropriate data governance and management practices*'. It follows that data governance and proper data management are essential for ensuring data availability and quality: large volumes of data are required for training AI, and data of sufficient quality is equally necessary to achieve greater accuracy and reliability in the outputs produced by AI systems.

In this regard, the technical report UNI CEI CEN/CLC/TR 18115:2025 '*Data Governance and Quality for Artificial Intelligence in the European Context*' provides a comprehensive overview of AI-related standards, with a particular focus on the data that powers artificial intelligence systems and their life cycles, ranging from data governance in individual projects to the level of global governance in the context of borderless phenomena such as climate, energy or pandemic risks.

In recognition of the importance of data, including for the creation of value, the European Union has for some years now established a legal and regulatory framework designed to facilitate the flow of data within the EU and across all sectors with a view to benefiting businesses, researchers and public administrations, while ensuring protection, fundamental rights, security and cybersecurity, all of which are aspects that must be taken into account when discussing AI. This framework includes, among other things:

- the adoption of a European strategy for data and the identification of common data spaces;
- the establishment of a data governance framework (through the Data Governance Act);
- the regulation of access to all data generated by connected products and related services (with the Data Act);
- the availability of public sector data for reuse (under the Open Data Directive).

[48] 'High-quality data and access to high-quality data plays a vital role in providing structure and in ensuring the performance of many AI systems, especially when techniques involving the training of models are used, with a view to ensure that the high-risk AI system performs as intended and safely and it does not become a source of discrimination prohibited by Union law'.

The issue of data management, at every stage of the process relating to the adoption, acquisition and development of AI solutions (training, grounding documents, outputs, etc.) in the public sector cannot be separated from the necessary coordination with existing policies and strategies at EU (as outlined above) and national level regarding data and open data; furthermore, the proper management of the data produced, from both a technical and contractual perspective, must be accompanied by the ability to acquire and reuse third-party datasets, obtained through various means, in order to preserve the dataset's reusability chain.

In this regard, the interoperability rules – particularly those relating to semantics – defined in accordance with the aforementioned regulatory framework, as well as the licensing strategies relating to the acquisition, processing, release/publication and reuse of public datasets, are of particular importance.

Without prejudice to the European Commission's ability to develop initiatives, including of a sectoral nature, to facilitate the lowering of technical barriers hindering cross-border exchange of data for AI development, including on data access infrastructure, semantic and technical interoperability of different types of data (see Recital (165) of the AI Act), public administrations MUST follow the guidelines already set out in guidelines and other binding documents, at both EU and national level, in terms of interoperability.

The fundamental requirement is that set out in the Guidelines for technical interoperability, which state that '*communication between entities MUST use shared data models in order to streamline and standardise the representation of information as a prerequisite for facilitating interoperability between different entities*'. In light of this, the rule that must generally take precedence is that (i) the data must be available through interoperable systems and that (ii) reference must be made to existing data models, ontologies and controlled vocabularies to verify whether the concepts already have widely adopted entities, properties and, where present, URIs, especially if in the European context. An example is provided by the interoperability framework relating to the spatial data domain defined by the so-called INSPIRE Directive (Directive 2007/2/EC) and the consequent implementing regulations on metadata and technical specifications of data and services, which is also indicated in Implementing Regulation (EU) 2023/138 as a reference for many of the high-value datasets. Legislative Decree No 36/2006 states in Article 6(9)

that, in the case of spatial data and environmental monitoring, the technical rules laid down in the INSPIRE Directive shall apply.

With regard to licensing strategies, the standardisation of licences adopted and to be adopted by the public sector introduced, in regulatory terms, by the 'Guidelines setting out technical rules for the opening up of data and the reuse of public sector information' ('the Open Data Guidelines')[49] becomes an absolute necessity, avoiding fragmentation – particularly for high-value data as referred to in Implementing Regulation (EU) 2023/138.

Furthermore, today, with the widespread adoption of computational analysis and training, the choice of licence should be guided, just as the production of technically high-quality data is, by the aim of ensuring optimal reusability and interpretability in the most transparent manner, including for new artificial intelligence systems that are already available or currently under development. It is strongly recommended that institutional datasets be used for training AI systems, since they ensure the reliability of the source and the robustness of the system that feeds the data.

The contractual clauses already envisaged, for example, in the operational guide on high-value data, which govern the acquisition of datasets produced by third parties, including in cases where such activities are outsourced, must also include provisions allowing for the reuse of such data for training or grounding purposes, or for the production of outputs from AI systems.

9.1. Data types

As mentioned in the introduction, data is the most significant and, at the same time, the most critical aspect of AI adoption in public administrations and beyond, forming the basis for the operation of any AI project.

The same data produced by the public administration may be used for multiple purposes and shared with multiple internal or external users; similarly, the public administration may use data from sources outside the organisation.

The data to be used and/or already used for AI can be characterised according to different parameters considered for the analysis. Considering the ever-changing

[49] Adopted by AgID by Decision of the Director-General No 183/2023 pursuant to Article 12 of Legislative Decree No 36/2006.

scenario, for example in relation to machine-learning algorithms that could give rise to other specific characteristics for information, it is not possible to produce a complete and definitive catalogue of data types for AI systems. The classifications set out below therefore represent a summary of the types that have currently been identified and may be subject to additions or amendments (e.g. the trend towards defining standardised data models will, hopefully, lead to the reduction or removal of unstructured data). In summary, technology and society are both changing, so the knowledge base on which algorithms are trained also needs to be updated over time.

For the purpose of building the model and deploying the AI system, the AI Act identifies:

- **training data**, defined as data used for training an AI system through fitting its learnable parameters (see Article 3(1)(29));
- **validation data**, defined as data used for providing an evaluation of the trained AI system and for tuning its non-learnable parameters and its learning process in order, inter alia, to prevent underfitting or overfitting (see Article 3(1)(30));
- **testing data**, defined as data used for providing an independent evaluation of the AI system in order to confirm the expected performance of that system before its placing on the market or putting into service (see Article 3(1)(32)).

With reference to the data sources, the following may be identified:

- **internal data** belonging to the public administration, i.e. native data generated by public administrations in the course of their institutional operations, data resulting from collaborations with other entities or data derived from integration with other data sources;
- **external data** from outside the public administration, acquired through procurement procedures, made available as open data by other organisations, shared in accordance with specific regulations (e.g. the Data Governance Act) or derived from other platforms (social media, IoT devices, etc.).

Considering the structure of the data, the following can be identified.

- **Structured data**, organised according to rigid, interrelated schemas and tables, potentially defined by specific data models (e.g. for geographical

data, the 'Content Specifications for Geotopographic Databases') and adhering to technical standards established at EU and national level (e.g. the guidelines on open, machine-readable formats set out in the Open Data Guidelines). In the case of such data, especially when based on shared data models, simplified data-cleaning operations may be provided for.

- **Unstructured data**, which, while containing a large amount of information, does not follow a specific reference model (e.g. images, videos, audio files, text documents, etc.). In such cases, data-cleaning operations are complex (reducing data volume, complexity and ambiguity, while improving accuracy, completeness and usability), given that pre-processing must also involve transforming the data into a format that can be used by AI algorithms.
- **Semi-structured data**, which contains information with hybrid characteristics (XML, JSON formats, etc.) and is characterised by certain organisational properties that facilitate analysis; in other words, it also contains additional information such as metadata or tags, making it more organised than unstructured data.

Depending on the stage of use/production of the data, the following may exist.

- **Input data**, defined by the AI Act as data provided to or directly acquired by an AI system on the basis of which the system produces an output (see Article 3(1)(33)). These may be training, validation or model-testing data (see above) or pre-processing data.
- **Output data**, which can be categorised as follows:
  - data generated by the model (text, images, audio);
  - performance data;
  - predictions and their associated probabilities and reliability.
- **Data from the AI system**, relating to the following aspects:
  - model parameters;
  - metadata;
  - source code of the model;
  - performance data;
  - log data;
  - status data;

- temporary data.

It should also be borne in mind that the use of AI techniques for predictive purposes is inevitably linked to the use of **historical time series data**, which represents a sequence of information relating to an event recorded at regular intervals over time. Historical time series data is an essential component for the implementation of statistical algorithms and machine-learning techniques aimed at the predictive analysis of a wide range of events. The AI Act points out that, especially in the case of the use of such data, biases may, for example, be inherent in underlying datasets.

Synthetic data artificially created for the training of AI systems, which is also referred to in the AI Act (see Recital (111), Article 10(5) and Article 59), must be taken into account. The effectiveness of synthetic data stems from the objective difficulty, in many situations, of collecting, structuring, processing and validating real data or complying with privacy requirements. This may slow down the implementation of AI systems. The use of synthetic data makes it possible to train and, where necessary, refine the algorithms of AI systems before real-world data becomes available. In relation to such data, key considerations may include effectively protecting personal data, ensuring that the data is representative of the real world, and thus that its statistical properties are appropriate, as also referred to in the AI Act, and ensuring the actual usefulness of such data in the implementation of AI systems.

Clearly, the above classifications of data should not be viewed in isolation; for example, input data may be structured, unstructured or semi-structured, and may be used for training, validation or testing.

Shared data spaces can play a crucial role in ensuring the continuous availability of data. These are set out in the European strategy for data, which has outlined a roadmap for their creation across a range of strategic sectors, with the aim of maximising the value of data for the benefit of the economies and social activities of Member States. These common spaces, which have broadened their scope over time and are an essential component of the single data market, can facilitate the sharing and pooling of reliable and secure data in strategic economic sectors and areas of public interest.

Currently, the Common European Data Spaces cover 14 sectors: agriculture, cultural heritage, energy, finance, the Green Deal, health, language,

manufacturing, media, mobility, public administration, research and innovation, skills, and tourism.

The first of the common spaces for which specific regulations have been introduced is that of health data, for which Regulation (EU) 2025/327 was recently adopted[50].

9.2. Data characteristics

As highlighted in the Italian AI Strategy 2016–2024, datasets are also becoming essential infrastructure across all aspects of developing and harnessing the new potential of AI within public administrations. It is therefore necessary to take coordinated, consistent action, prioritising the use of knowledge to create, technically and legally safeguard, and publish high-quality and genuinely reusable datasets.

As a preliminary step, public administrations MUST undertake a systematic analysis of the status and quality of their datasets, taking steps to update and publish, where this has not yet been done, data that can be published as open data, in accordance with the Open Data Guidelines and the Operational Guide on High-Value Datasets, releasing the data under the CC BY 4.0 licence (or, where possible, CC0) or another equally permissive licence, such as the CDLA Permissive 2.0, which already explicitly takes computational analysis into account. In addition to this, there are issues relating to the processing of personal data and intellectual property when using open data for the training, validation and testing of AI systems for which, in addition to the discussion in other sections of this document, the guidance set out in the Open Data Guidelines should be referred to.

This initiative is also in line with Principle 2.1 of the OECD's recommendations document[51], which urges governments to invest in open datasets that are representative and respect privacy and data protection to support an environment for AI research and development that is free of harmful bias and to improve interoperability and use of standards.

Furthermore, given the significant volume of datasets that cannot be published as open data (for example, because they contain personal data or data protected by intellectual property rights), it is essential to establish a secure,

[50] https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=OJ:L_202500327
[51] Recommendation of the Council on Artificial Intelligence

common framework to implement, as far as possible, the provisions of the Data Governance Act (Regulation (EU) 2022/868). In this regard, it will be essential to proceed in accordance with the provisions of Legislative Decree No 144/2024 by establishing support and guidance programmes to assist public bodies in defining the required technical and contractual tools.

AI systems learn and produce reliable analyses if they have access to a sufficient amount of high-quality data. Compliance with personal data protection and security regulations MUST also be ensured. These matters are addressed in specific sections of this document.

With regard to quality, the Open Data Guidelines provide specific indications referring to the UNI CEI ISO/IEC 25012:2014 and UNI CEI ISO/IEC 25024:2016 standards. These standards define a set of characteristics for quality (and related measures) divided into 'inherent' (accuracy, updating (currency), completeness, consistency (coherence), credibility), 'inherent and dependent on the system' (accessibility, comprehensibility, compliance, efficiency, accuracy, confidentiality, traceability) and 'system-dependent' (availability, portability and restorability).

With regard to these characteristics, the Open Data Guidelines, drawing on AgID's Commissioner's Decision No 68/2013, recommend that at least four data quality characteristics be ensured, namely accuracy, coherency, completeness and currency (see Section 5.3 of the Open Data Guidelines). Compliance with other quality criteria, which are also relevant to AI systems, is ensured through obligations arising from specific regulations, such as accessibility (as set out in Law No 4/2004 and the relevant AgID Guidelines) or confidentiality in accordance with the provisions of the GDPR.

The guidelines set out here also apply in the context of AI, given that the approach to be adopted is not to define a new model and new processes relating to data quality, but rather to integrate the elements, measures and procedures that are specific to AI.

In addition to the quality characteristics set out in the Open Data Guidelines (**accuracy**, **coherence**, **completeness** and **currency**) and those arising from specific regulatory obligations (**accessibility** and **confidentiality**), taking into account the quality requirements identified in the AI Act, AI systems MUST also guarantee the data quality characteristics listed below, which are derived from the

ISO/IEC 25012 and ISO/IEC 5259-2 standards, with additional requirements depending on the context.

1. **Representativeness:** the degree to which the datasets reflect the population being studied (ISO/IEC 5259-2).
2. **Balancing:** the distribution of samples across all features of the datasets (ISO/IEC 5259-2).
3. **Traceability:** the extent to which data has attributes that provide a record of access to the data and of all changes made to the data within a specific usage context (ISO/IEC 25012).
4. **Availability:** the degree to which data has attributes that allow it to be recalled by authorised users and/or applications in a specific use context. The data must be stored in an organised and structured manner, with adequate systems to quickly retrieve the information necessary for training and implementing algorithms (ISO/IEC 25012).
5. **Credibility:** the extent to which data possesses attributes that are considered true and credible by users in a specific use context (ISO/IEC 25012).

The amount of data is another important factor for effectively training AI systems. In particular, it is relevant in big data, defined in the ISO/IEC 20546:2019 standard as extensive datasets – primarily in the data characteristics of volume, variety, velocity, and/or variability – that require a scalable technology for efficient storage, manipulation, management, and analysis. For big data, the standards highlight four characteristics: two for structure and two for quality. The structural characteristics are the volume and variability of the sources. Large volumes of data and their variability may require the use of automated tools, as indicated in the ISO/IEC 5259-1 standard.

In short, when it comes to data volume, AI systems SHOULD be fed large amounts of data in order to train complex models. Similarly, the data MUST also include a variety of examples representative of the possible situations that the AI systems are assessing.

Another important aspect is 'provide plain and easy-to-understand information on the sources of data/input, factors, processes and/or logic that led to the prediction, content, recommendation or decision, to enable those affected by an AI system to understand the output' as stated in Principle 1.3 'Transparency and explainability' of the OECD Recommendations document mentioned above. In this

regard, the Italian AI Strategy 2024–2026 has set out a programme aimed at establishing a register of datasets and models that are built in accordance with principles of transparency and fairness, are ethically reliable by design, and are reusable to accelerate the development of solutions by Italian companies. According to the Strategy, all projects funded under the same national strategy or otherwise receiving public funding will be required to report the datasets used and the models produced in the register, in accordance with guidelines that will define the levels of access and methods for reuse.

In relation to this, it should be noted that a semantic model is being defined in the field of machine learning as part of the European Commission's SEMIC initiative, the objective of which is to extend the use of DCAT-AP (the metadata profile used for open data). The MLDCAT-AP model[52], currently at the 'Candidate Recommendation' stage, facilitates standardised descriptions of a machine-learning process, along with the associated datasets, the quality metrics applied to the datasets and citations of relevant documents.

In addition to the more conventional sense of data being put 'at the service' of AI, the reverse relationship is also significant, whereby AI acts as a tool for improving the quality of the data itself, as well as its representativeness, completeness and interoperability. In this regard, it is possible to harmonise classifications and descriptions of items of the same type, standardise, clean up and improve the accuracy of archives, extract key terms for subsequent use in searches, model and represent the topics covered by a text corpus and, in general, summarise and effectively manage a body of information which, by its very nature, is largely unstructured but which, through AI, can be organised and 'mastered'.

Another aspect in this context concerns the need and the possibility – including through the use of AI – to make data 'communicate' with other data, in other words, make it interoperable and integrated, thereby expanding the knowledge base and opening up significant opportunities for the data itself to regenerate and realise its full potential, in addition to what it can already achieve when its use is confined to the 'silos' within which it was produced or acquired.

[52] https://semiceu.github.io/MLDCAT-AP/releases/2.0.0/

9.3. Data processes and governance

To ensure proper data management when implementing AI solutions in public administrations, a structured and accountable process MUST be established that covers the entire data life cycle, ensuring that collection, storage, processing, analysis, monitoring and updating are secure and compliant with current regulations. It is also necessary to provide guidance to ensure data quality and availability, personal data protection and data security, resilience against bias, the adoption of ethical and transparent practices and continuous compliance with evolving regulations.

Furthermore, the actions of public administrations cannot disregard (or deviate from) a set of values, principles, standards and requirements such as openness, transparency, accountability and the pursuit of greater efficiency, must not compromise effectiveness (for example, by increasing inequalities), and must include an assessment of overall cost effectiveness, taking environmental impacts into account.

This should not, however, prefigure the definition of new data management processes for AI that are separate from the procedures followed (so far) within the organisation. Instead, a broader, structured data management process MUST be clearly defined (if it has not already been codified and implemented) to oversee the entire data life cycle, regardless of its use, and which incorporates procedures and practices specific to the context of AI. Data management for the purposes of AI systems, therefore, can be considered a subset of the overall data management.

The Open Data Guidelines include a section on organisational aspects in which a data preparation path is defined (see Figure 2 of the same Guidelines) as the result of a chain of processes and a series of analysis and processing activities aimed at improving the quality of and access to the data itself, which are then the basic requirements required for AI systems.

In line with the approach outlined above, which considers a general data management process while also taking into account aspects and characteristics relating to AI, the framework set out in the aforementioned Open Data Guidelines can be appropriately revised and adapted to include specific stages and activities for AI. For this purpose, although it refers to high-risk AI systems, the aspects

mentioned in Article 10 of the AI Act relating to data governance and management practices may be considered, namely:

a) the relevant design choices;

b) data collection processes and the origin of data, and in the case of personal data, the original purpose of the data collection;

c) relevant data-preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation;

d) the formulation of assumptions, in particular with respect to the information that the data are supposed to measure and represent;

e) an assessment of the availability, quantity and suitability of the datasets that are needed;

f) an examination in view of possible biases that are likely to affect the health and safety of persons.

Some of the aspects mentioned are already covered, or may be covered, in the various stages of the process referred to above. The activities referred to in paragraphs (b) and (d), for example, may be considered at the stages referred to as 'identification' and 'analysis', and those referred to in paragraphs (c) at the stages referred to as 'analysis' and 'enrichment'.

Similarly, the data life cycle described in the ISO/IEC 8183 standard may be considered. This standard identifies 10 stages (with specific details provided for each):

- **Stage 1**. Idea conception (business objectives and metrics);
- **Stage 2.** Business requirements (objectives, strategy, business and user requirements, compliance);
- **Stage 3**. Data planning (volume of data required, source, synthetic data, format, security, personal data protection);
- **Stage 4**. Data acquisition (from internal sources, third parties, open data);
- **Stage 5**. Data preparation (cleaning, processing, standardisation, data organisation, labelling, resampling, coding, integrity verification, provenance, anonymisation or pseudonymisation of data);
- **Stage 6**. Building model (training of an ML algorithm, combination of human knowledge);

- **Stage 7**. System deployment (the AI system goes live in the target environment);
- **Stage 8**. System operation (data analysis, data visualisation, data transmission, data storage);
- **Stage 9**. Data decommissioning (secure deletion, archiving, reuse, retention for audit purposes);
- **Stage 10**. System decommissioning (cessation of data processing, storage of logs).

Here too, some of the stages outlined overlap or can otherwise be incorporated into the process set out in the Open Data Guidelines: the activities specified in stages 3 and 4 of ISO/IEC 8183 can be considered within the stages of the guidelines identified as 'identification' and 'analysis'; those set out in stage 5 of ISO/IEC 8183 can be considered in the stages of the guidelines identified as 'analysis', 'enrichment', 'modelling and documentation' and 'validation'.

It follows that a model for an overall data management process – adapted, as mentioned, from the Open Data Guidelines and incorporating elements from the AI Act and the ISO standards cited – could be as shown in Figure 5. The stages with a grey background are those relating to activities and operations aimed exclusively at AI.

Figure 5. Data management process model.

Identificazione
Pianificazione
Raccolta
Analisi dei vincoli
Analisi
Analisi della qualità
Pulizia e bonifica
Analisi di processo

Identification
Planning
Collection
Constraint analysis
Analysis
Quality analysis
Cleaning and remediation
Process analysis

Preparazione	**Preparation**
Arricchimento	Enrichment
Modellazione e documentaz.	Modelling and documentation
Trasformazione aggregazione campionamento etichettatura anonimizzazione	Transformation, aggregation, sampling, labelling, anonymisation
Validazione	**Validation**
Valutazione della qualità	Quality assessment
Verifica dell'integrità	Integrity verification
Valutazione di possibili distorsioni	Assessment of possible biases
Utilizzazione	**Use**
Pubblicazione e/o disponibilità	Publication and/or availability
Archiviazione e conservazione	Archiving and storage
Addestramento implementazione e funzionamento del sistema	Training, implementation and operation of the system
Monitoraggio	**Monitoring**
Revisione e aggiornamento	Review and updating
Dismissione dei dati	Data decommissioning

As is clear, several sub-stages are general in nature and can therefore be applied for a variety of purposes, not just for AI, depending on the processes in question (open data publication, management of protected data, adoption of AI systems, etc.).

Below is a more detailed look at the stages and activities outlined in the diagram above.

Identification stage:

Planning is related to the definition of a data collection and management strategy that ensures the quality, variety and relevance of the data for the specific objective that may be pursued. The objective is, therefore, to identify the data needs necessary for the various purposes and to plan the activities necessary for their use (e.g. with regard to security). In the case of data management for AI, this includes: amount of data needed, source, need for synthetic data (see ISO/IEC 8183, stage 3). Prior to this (see ISO/IEC 8183 stage 2) is, inter alia, the identification of the most appropriate data architecture (e.g. Data Lake, Data Mesh or Data Fabric) and the resulting technology, while also making reference to innovative approaches for data management.

The data collection process involves both the identification of data (already covered in the Open Data Guidelines) derived from internal sources, that is, data produced within the administration or managed by it (for example, resulting from mashup operations), and the acquisition of data from other sources, such as APIs, IoT sensors, social media, images and public/open data from other public administrations (see ISO/IEC 8183 stage 4).

The purpose of the constraint analysis is to identify the barriers that prevent data from being used for a particular purpose. In the case of AI, for example, this means verification of the original purpose of the data collection where personal data is concerned.

Analysis stage:

The purpose of quality analysis is to assess the initial quality of the collected data so that appropriate data cleaning and remediation can be carried out.

The cleaning/remediation process is aimed at improving data quality through the removal or correction of duplicates, errors and missing values using data-based processes or adopting process-based remediation actions (see ISO/IEC 8183 stage 5).

As indicated in the Open Data Guidelines, the analysis of the organisational process that produces and manages the data is necessary in order to ensure that the production of that data is consolidated and becomes stable, according to the update frequency and the release methods adopted.

Preparation stage:

The preparation stage includes the stages identified as 'Enrichment' and 'Modelling and Documentation' (relating to data standardisation and organisation) as set out in the Open Data Guidelines (to which reference should be made for further detail), with the addition of certain specific pre-processing operations designed to make the data suitable for training AI models. Such operations may include transformation, resampling, labelling, coding, data masking and the anonymisation or pseudonymisation of data (see ISO/IEC 8183 stage 5).

Validation stage:

The quality assessment process, as set out in the Open Data Guidelines, includes integrity checks (see ISO/IEC 8183 stage 5) and assessments of potential biases, one of the aspects highlighted by the AI Act. Additional specific checks, to be conducted through periodic audits, may be required in individual domains.

Use stage:

The use stage includes publication activities aimed at ensuring the availability of data and data storage (which must be carried out in a secure and scalable manner, typically using cloud infrastructure that guarantees elastic scalability,

distributed storage, data security and data encryption); stages 6, 7 and 8 as set out in the ISO/IEC 8183 standard are included in this.

Monitoring stage:

The constant monitoring of the quality and freshness of data which, by its very nature, is not static, particularly when working in AI contexts, must be accompanied by periodic reviews, updates and audits to take account of any changes in the domain, in the behaviours observed or in the relevant data sources. Consideration must also be given to data decommissioning procedures (see ISO/IEC 8183 stage 9) in cases where data is no longer used by the system (e.g. secure deletion, archiving, reprocessing). The standard states, among other things, that at this stage certain categories of data should be retained for auditing purposes (e.g. log data to demonstrate compliance) and that data may be deleted in accordance with explicit licence terms or in compliance with data protection legislation.

In summary, public administrations MUST ensure effective and robust data life cycle management, with particular attention paid to the following processes:

- **Collection and Storage**: The data MUST be collected and stored in a secure manner and in accordance with the regulations in force.
- **Processing and Analysis**: The data MUST be processed and analysed using advanced techniques to ensure its integrity and quality.
- **Monitoring and Updating**: Datasets MUST be updated and monitored through governance processes to ensure that analyses remain relevant and accurate.

Figure 6 illustrates the correlation between the stages of the AI life cycle and the stages of the data management process. Without prejudice to the fact that the data quality assessment must continue throughout the entire life cycle of the data, it is also proposed here, similarly to the ISO/IEC 5259-1:2024 standard, when during the life cycle the quality characteristics considered in the model defined in this document should be measured.

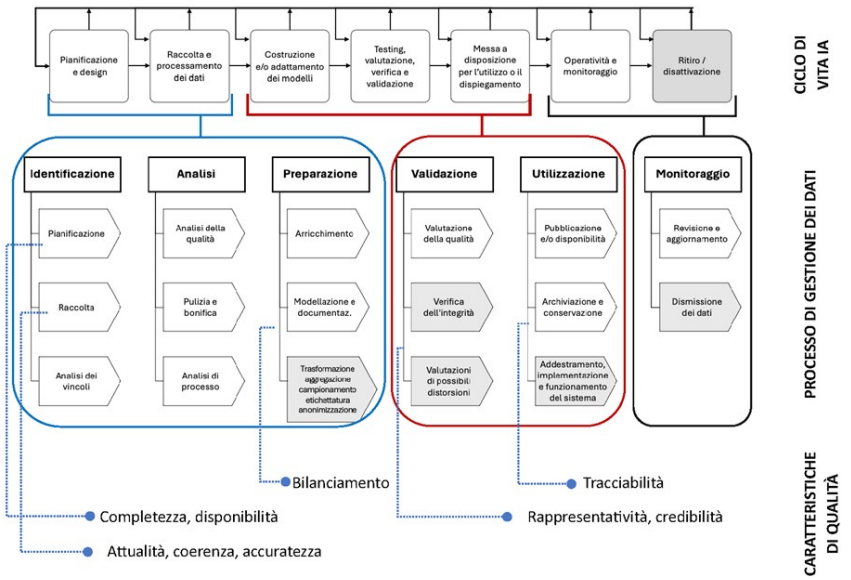


Figure 6. Correlation between the stages of the AI life cycle and the stages of the data management process.

CARATTERISTICHE DI QUALITA PROCESSO DI GESTIONE DEI DATI	**QUALITY CHARACTERISTICS DATA MANAGEMENT PROCESS**
CICLO DI VITA IA	**AI LIFE CYCLE**
Pianificazione e design	Planning & design
Raccolta e processamento dei dati	Data collection and processing
Costruzione e/o adattamento dei modelli	Build and/or adapt model(s)
Testing, valutazione, verifica e validazione	Test, evaluate, verify & validate
Messa a disposizione o il dispiegamento	Make available for use or deploy
Operatività e monitoraggio	Operate and monitor
Ritiro/disattivazione	Retire / decommission
Completezza, disponibilità	Completeness, availability
Attualità, coerenza, accuratezza	Currency, consistency, accuracy
Bilanciamento	Balancing
Tracciabilità	Traceability
Rappresentatività, credibilità	Representativeness, credibility

Data management, even in the context of adopting AI solutions, cannot be separated from the proper establishment of data governance, with a clear definition of roles and associated responsibilities. This aspect is also addressed, in relation to that area, in the Open Data Guidelines, which refer to Circular No 3 of 1 October 2018 issued by the Minister for Public Administration, recommending that the Digital Transition Manager (DTM) be granted the power to set up thematic groups for specific activities and/or obligations. In the process of opening up and publishing data, it is therefore recommended that a dedicated working group be set up within the organisation to oversee the data-opening process, establishing, where possible, the appropriate bodies and roles required for this purpose, taking

into account the points of contact for each individual domain and ensuring the necessary involvement of the Data Protection Officer where personal data is involved.

The establishment of specific coordination bodies or working groups – whether existing or to be set up – within the DTM's office is also provided for in the Directive issued by the State Secretary, Prime Minister's Office, responsible for matters relating to technological innovation and digital transition, entitled '*Measures for the implementation of Article 50-ter of Legislative Decree No 82 of 7 March 2005*'[53]. According to the Directive, the tasks to be entrusted to these bodies also include achieving data governance objectives and streamlining existing databases within the public administration in order to ensure the uniqueness and quality of data, and promoting the sharing of information held by public administrations.

With a view to preventing the proliferation of organisational bodies dedicated to data, data governance within public administrations, including in relation to data for AI, SHOULD be brought under a single coordinating body responsible for overseeing their overall management as outlined above (and which therefore covers all the aspects mentioned: open data, implementation of Article 50-ter, data for AI, etc.).

In this context, the issue of institutional collaboration, which is also addressed in the three-year plan, is also relevant. In line with the provisions of that plan, certain public administrations may also play a coordinating role in the data management process (as national and/or regional and/or provincial hubs), acting as intermediaries, implementers and promoters of solutions for local bodies, listening to their needs and complementing the varying priorities and capabilities expressed by different local public administrations.

These authorities act as intermediaries and liaise with lower-level local authorities, leveraging the technological capabilities, spending power and cost-effectiveness afforded by their larger geographical scope. In the field of data management, too, the role played by higher-level authorities is considered crucial as central coordinators of infrastructure, services and contracts for the collection and management of an integrated public information resource to support AI, carried out and directly overseen by each local authority. This approach ensures a

[53] https://www.governo.it/sites/governo.it/files/Decreto20231205_Direttiva_PDND.pdf

wider range of solutions, consistent and uniform management and economies of scale in the data collection and pre-processing stages.

This can be translated into the promotion of local initiatives for leveraging data, such as the creation of data centres or interregional high-performance computing solutions, the promotion of open data hub platforms, the development of partnerships with local bodies, universities, research centres and companies for data analysis and use projects, investment in training programmes and skills development in the data sector for civil servants and citizens.

Initial mapping of the maturity level and the needs of the various local administrations in relation to the above-mentioned areas is certainly useful in facilitating consistent development across these areas, by identifying scenarios and strategies that the higher level of governance can pursue and recommend to the local administrations, or implement directly where appropriate. Such mapping is also useful for fostering an awareness of, and a willingness to adopt, a strategy for managing and leveraging data, which in certain situations is inevitably collected for purely administrative purposes but is then either not used at all or used in silos, in line with innovative approaches driven by a 'bottom-up' approach from a small number of individuals who have personally recognised the benefits.

The involvement of pioneering administrations, in relation to the circumstances of the bodies operating in the local area, may ultimately involve a number of aspects, including the following.

- The centralised implementation and management of infrastructure and tools, based on a 'cloud'-style approach, to be made available to each public body, pooling a range of investments and skills and achieving clear economies of scale (also significant in this context is the possibility of federating computational resources across central, regional and interregional levels, as a potential avenue for sharing the computing resources of Italian public administrations, in a spirit of institutional cooperation).
- The very availability of data, which is often of broad interest, and the collecting and sharing of that data, with the higher level therefore acting as an aggregator and coordinator for the benefit of the entire territory, ensuring consistent quality standards and updates, and thereby generating significant cost savings.

- The provision, in its capacity as an aggregator, of contractual tools designed to carry out projects to leverage and reuse data. For each local body, taken individually, it can be difficult, burdensome and ineffective to activate public procedures for the selection of providers offering this type of innovative service, potentially giving rise to uncoordinated initiatives.
- The launch of initiatives to encourage public bodies to adopt new methods for managing, processing and leveraging public information assets, which may incentivise their active use and participation and which may be based on a combination of financial contributions, technical support, training and awareness raising.
- The prospect of synergies between bodies, which may enter into agreements to share the facilities, human resources and skills required to carry out projects of mutual interest.

It is assumed that initiatives such as those outlined above will encourage local bodies to adopt more structured and advanced practices for managing and leveraging data, helping administrations to become more efficient, transparent and capable of offering additional services within their respective areas of responsibility, viewing each administration as part of a 'digital administrative ecosystem', within a collaborative network of public bodies, private companies and citizens, aimed at sharing and leveraging data, which can lead to innovative solutions in various sectors, such as transport, energy or environmental management.

10. Personal data protection

In the context of these Guidelines, AI systems may be used to carry out and/or support activities involving the processing of personal data: the adoption of such systems by public administrations MUST therefore take place in accordance with the fundamental right to the protection of personal data.

With this in mind, when adopting an AI system, the public administration MUST give primary consideration to assessing and verifying the AI system's compliance with EU and national data protection legislation, as well as to the impact the AI system will have on the rights of the data subjects concerned, including by adapting its organisational model and the related technical and organisational security measures.

The AI Act itself makes it clear that EU law on the protection of personal data, privacy and the confidentiality of communications applies to personal data processed in connection with the rights and obligations established by the AI Act, and further specifies that the application of existing EU law on the processing of personal data is not affected, including the tasks and powers of the independent supervisory authorities responsible for monitoring compliance with those instruments, and that the obligations of providers and deployers of AI systems in their role as controllers or processors of personal data are likewise unaffected.

In particular, the public administration MUST ensure compliance with the principles set out in Article 5 of the GDPR, including the accountability of the public administration itself when acting as a data controller, in accordance with EU and national legislation on the protection of personal data and the decisions and opinions issued by the European Data Protection Supervisor, the European Data Protection Board and the Italian Data Protection Authority.

Where an AI system is used to carry out activities involving the processing of personal data, the public administration MUST conduct an analysis of that system from the specific perspective of personal data protection in order to ensure and demonstrate compliance with the following[54]:

[54] For further details on the issues discussed, please refer to the guidelines issued by the European Data Protection Board and the former Article 29 Working Party, as well as those issued by the Italian Data Protection Authority.
These include, for example, the '*Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*', adopted by the Article 29 Working Party on 3 October 2017, as amended and adopted on 6 February 2018, available at: https://ec.europa.eu/newsroom/article29/items/612053/en.

- AI systems processing personal data do so in a lawful, fair and transparent manner vis-à-vis the data subject; in particular, the PA MUST assess and identify the necessary grounds for lawfulness and the necessary guarantees of transparency;
- personal data is collected via the AI system for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; in particular, when using AI tools in the performance of its institutional duties, the PA MUST refer to the specific purposes of public interest, within the scope of the powers conferred upon it by law;
- a suitable legal basis – including that provided by Law No 132 of 23 September 2025 – is always identified for the processing of personal data using the AI system, from the very outset of its development or acquisition, including with regard to any subsequent use of the data for purposes other than those for which it was collected;
- the personal data processed using the AI system must be adequate, relevant and limited to what is necessary in relation to the purposes for which it is processed; in particular, the PA MUST assess whether there is a genuine need to use personal data and, if so, ensure that such data is minimised;
- the personal data processed using the AI system is accurate and, where necessary, kept up to date, with the public administration taking every reasonable step to ensure that any data that is inaccurate or out of date in relation to the purposes for which it is processed is erased or rectified without delay;
- the personal data processed using the AI system is kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data is processed, and measures concerning the storage methods and retention period for personal data processed using AI systems are appropriately identified and established;
- the PA uses the AI system in such a way as to ensure a level of personal data security appropriate to the risk, by identifying and implementing organisational and technical measures including those specifically related to the use of AI tools, that are adequate to protect data from security breaches which could result, unlawfully or accidentally, in the loss, alteration,

unauthorised disclosure or unauthorised access to personal data that is transmitted, stored or otherwise processed;

- the PA always guarantees that data subjects may exercise their rights regarding the protection of personal data;
- the PA instructs and designates the staff to whom it assigns specific tasks and functions related to the processing of personal data by means of the AI system that has been adopted;
- where the PA intends to determine the purposes and means of processing jointly with another entity or to use another entity to process personal data using an AI system, it always acts in accordance with Articles 26 and 28 of the GDPR;
- when using the AI system, the PA ensures the protection of personal data by design and by default;
- before processing personal data using the AI system that has been adopted, and subsequently on a regular basis – including where AI systems are adopted that do not qualify as high risk and are therefore not subject to the obligation to carry out an assessment of the impact on fundamental rights under Article 27 of the AI Act – the PA carries out a data protection impact assessment in accordance with Article 35 of the GDPR or Article 27 of Directive (EU) 2016/680, the Guidelines of the Article 29 Working Party[55] and the provisions of the Data Protection Authority, identifying the risks and the appropriate technical and organisational measures to mitigate them and, where a high risk is identified in the absence of mitigation measures, consults the Data Protection Authority; the extent of the adverse impact of the AI system on fundamental rights, including the protection of personal data, is indeed of particular relevance for the purposes of classifying an AI system as high risk;
- special categories of personal data and personal data concerning criminal convictions and offences, as identified in Articles 9 and 10 of the GDPR, are processed by means of AI systems subject to the identification of a suitable legal basis and appropriate, specific and documented technical and organisational security measures in compliance with the principle of

[55] See the '*Guidelines on Data Protection Impact Assessments (DPIAs) and determining whether processing is 'likely to result in a high risk' for the purposes of Regulation (EU) 2016/679*' of the Article 29 Working Party of 4 April 2017, as last amended and adopted on 4 October 2017 and adopted by the European Data Protection Board on 25 May 2018, available at: https://ec.europa.eu/newsroom/article29/items/611236/en.

accountability, not only in the necessary compliance with the provisions of EU and national legislation on the protection of personal data but also under the AI Act itself, which provides specific rules on the protection of natural persons with regard to the processing of personal data, consisting, inter alia, of restrictions on the use of AI systems for remote biometric identification for law enforcement purposes, the use of AI systems for risk assessment of natural persons for law enforcement purposes and the use of biometric categorisation AI systems for law enforcement purposes, recalling for such contexts the provisions of Article 16 of the Treaty on the Functioning of the European Union.

In the context covered by these Guidelines, however, it is essential that public administrations adopting AI systems to carry out their activities are fully aware that they act as data controllers within the meaning of Article 4(7) of the GDPR and, therefore, that they bear responsibility – including in terms of accountability, pursuant to Article 5(2) of the GDPR – for the processing carried out via the AI tool that has been adopted, even in cases where such use is facilitated through the relevant provider.

In this context, the public administration MUST highlight the role and involvement of the Data Protection Officer (DPO), in compliance with the provisions of the GDPR regarding the position and tasks of the DPO, including by identifying forms of collaboration with the Digital Transition Manager.

Furthermore, as highlighted by the Data Protection Authority, the public administration MUST pay the utmost attention to three key principles that must necessarily govern the use of algorithms and AI systems in the performance of tasks of significant public interest:

'1. the principle of knowability, whereby the data subject has the right to know about the existence of decision-making processes based on automated processing and, where this is the case, to receive meaningful information on the logic used, so as to be able to understand it;

2. the principle of non-exclusivity of algorithmic decision-making, according to which there must always be human intervention in the decision-making process capable of monitoring, validating or overturning the automated decision (human-in-the-loop);

3. the principle of algorithmic non-discrimination, according to which the data controller should use reliable AI systems that minimise opacity and errors caused by technological and/or human factors, periodically reviewing their effectiveness in light of the rapid evolution of the technologies used and the appropriate mathematical or statistical procedures for profiling, while implementing appropriate technical and organisational measures. This is also to ensure that factors leading to data inaccuracies are rectified and that the risk of errors is minimised, given the potential discriminatory effects that inaccurate processing of health data may have on natural persons (see Recital (71) of the Regulation).[56]

The European Data Protection Board has also recently commented on certain data protection aspects related to the processing of personal data in the context of AI models in its Opinion No 28 of 17 December 2024, which serves as a useful guide for public administrations regarding the nature of AI models in relation to the definition of personal data, the circumstances in which AI models might be considered anonymous and the related demonstration thereof, the adequacy of legitimate interest as a legal basis for the processing of personal data in the context of the development and implementation of AI models and the possible impact of the unlawful processing of personal data during the development of an AI model on the lawfulness of the subsequent processing or operation of the AI model[57].

In particular, in light of the points highlighted in the aforementioned Opinion No 28 of 17 December 2024 of the European Data Protection Board, the public administration MUST assess whether or not the AI model can be considered anonymous, with particular reference to the likelihood of obtaining or reconstructing, intentionally or otherwise, the personal data used to train the model, including through the mere use of the AI system.

[56] See '*A Guide to the Implementation of National Healthcare Services Using Artificial Intelligence Systems*' at https://www.garanteprivacy.it/documents/10160/0/Decalogo+per+la+realizzazione+di+servizi+sanitari+nazionali+attraverso+sistemi+di+Intelligenza+Artificiale.pdf/a5c4a24d-4823-e014-93bf-1543f1331670?version=2.0.
It should be clarified that, although the aforementioned measure is specifically aimed at the healthcare sector, the key principles outlined above are considered essential for the informed and appropriate management of AI algorithms and systems when carrying out tasks of significant public interest.

[57] See EDPB Opinion No 24/2024, available at: https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en.

Furthermore, with reference to the above, where public administrations make use of data processors that use AI tools, the public administration MUST explicitly state including through the designation referred to in Article 28 of the GDPR, that personal data must not be processed for the purpose of training or retraining models to be made available to third parties, in accordance with the principle of purpose limitation and the obligations incumbent upon data processors.

The European Data Protection Board has also recently presented two new projects from the Support Pool of Experts: one on the topic of 'Law & compliance in AI security and data protection' and the other on the topic of 'Fundamentals of secure AI systems with personal data'. The two projects, initiated at the request of the Hellenic Data Protection Authority (HDPA), provide training materials on AI and data protection. The documents produced could be a useful guidance tool for data protection experts and PAs in the adoption and use of AI systems.

11. Cybersecurity

The security of AI systems is an essential requirement for the adoption, procurement and development of AI within public administrations. While AI appears to be able to provide useful tools to respond to the increasing need to improve efficiency and effectiveness in the management and delivery of public services[58], this technology also introduces new risks, threats and vulnerabilities that must be taken into account in its implementation.

The security of an AI system can be defined[59] as the tools, strategies, and processes implemented that identify and prevent threats and attacks that could compromise the confidentiality, integrity, or availability of an AI model or AI-enabled system.

In view of their inherent **socio-technical** nature – in which social elements (the influence of social dynamics and the impact on people who use or are affected by them) and technical elements (such as datasets, algorithms and models) are closely intertwined – AI systems are characterised by specific security features, in particular with regard to the *risks* to which they are subject and the *attacks* to which they are subjected.

In addition to the valid references presented in this chapter (e.g. NIST, ENISA), PAs MUST take into account the technical standards produced by CEN-CENELEC JTC21 'Artificial Intelligence', following the European Commission's request for standardisation (see Section 4.4.), as they represent harmonised technical standards capable of ensuring the presumption of conformity with the cybersecurity requirements laid down in Article 15 of the AI Act.

11.1. Attack taxonomies

In addition to traditional cyber[60] attacks on the underlying ICT infrastructure, AI systems are vulnerable to attacks targeting specific AI components, such as models and training data. Understanding the various attack

[58] Agency for Digital Italy (AgID), Three-Year Plan for Information Technology in Public Administration 2024–2026 edition, 2025 update, https://www.agid.gov.it/it/agenzia/piano-triennale.

[59] The definition is taken from the article published on the MITRE ATLAS website at https://atlas.mitre.org/resources/ai-security-101. MITRE ATLAS is a knowledge base that documents adversarial tactics and techniques relating to AI systems.

[60] Attacks in which the malicious actor makes use of TTPs (tactics, techniques and procedures that characterise an attacker's behaviour to achieve his or her objective) typical of the cyber domain.

taxonomies allows targeted and contextualised protection and containment actions to be implemented.

The taxonomy developed by NIST[61] may be used as a reference framework for attacks on AI systems. It identifies the following broad categories of attacks:

- *evasion attacks;*
- *poisoning attacks;*
- *privacy attacks;*
- abuse attacks;

The first three categories cover both predictive and generative AI models, while the last category covers only generative AI models[62].

These categories and possible mitigation strategies are briefly discussed in the following sections.

11.1.1. Evasion attacks

This category of attacks aims to generate an error in the classification of the model by introducing perturbations (often imperceptible to humans) in the model's *inputs,* called *adversarial examples.*

These attacks are designed to exploit vulnerabilities in the model's decision-making process. They may cause the model to predict a value desired by the attacker or result in a reduction in the accuracy of the model.[63]

Possible mitigation strategies include**, adversarial training** (involving retraining the model using correctly labelled adversarial examples), **randomised smoothing** and **formal verification**. These approaches aim to make the model invariant to any noise introduced by an attacker, thereby increasing its robustness.

11.1.2. Poisoning attacks

This category of attacks aims to degrade a model's performance or cause it to produce a specific output by tampering with (*or poisoning*) the model's training data. An example of an attack is 'label flipping', in which the attacker changes the

[61] National Institute of Standards and Technology (NIST), Adversarial Machine Learning – Taxonomy and Terminology of Attacks and Mitigations, 2024,

[62] Predictive AI systems identify patterns and relationships or make predictions based on training data, while generative AI systems create new content based on training data.

[63] For example, in *K. Eykholt et al., 'Robust Physical-World Attacks on Deep Learning Visual Classification,' 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 1625–1634, doi: 10.1109/CVPR.2018.00175* it has been demonstrated that, through this type of attack, it is possible to mislead a model trained to recognise road signs, such as stop signs, by applying small stickers to them.

label of the training data with the aim of training the model on the basis of the label of the attacker's choice.[64]

These attacks can be categorised as follows.

- **Availability poisoning**: these cause a breach of the model's availability by degrading its performance across all sample data. These can be identified by monitoring the model's performance; possible mitigation strategies include cleaning the training data and robust learning (e.g. multiple models can be trained).
- **Targeted poisoning**: these compromise the model's integrity by altering its predictions on a small number of targeted samples. Possible mitigation strategies include implementing security controls to verify the origin and integrity of data.
- **Backdoor poisoning**: similarly to targeted poisoning attacks, this results in a breach of the model's integrity; in this case, however, the aim is to mislead the model in response to a specific sample of data (referred to as a trigger). Possible mitigation strategies include cleaning the training data, reconstructing the trigger and inspecting and cleaning the model.
- **Model poisoning:** directly modifying the trained model by injecting it with malicious functionality. These attacks can compromise both the integrity and availability of the system and generally occur within the context of federated learning, in which client systems send local model updates to a server, which aggregates them into a global model. They are also possible in supply chain scenarios in which models or related components that have been poisoned are acquired. Possible mitigation strategies include identifying and blocking malicious updates or (in the case of backdoor model poisoning) inspecting and cleaning the model.

11.1.3. Privacy attacks

This category of attacks aims to compromise user information by *reconstructing it* from training data. These attacks can be divided into:

- **data reconstruction**, reconstructing information from aggregated data;
- **membership inference**, determining whether a particular record has been included in the dataset used to train a model, thereby compromising user information;

[64] For example, when training an email spam filter, an attacker could change the labels of the training data from 'spam' to 'not spam' in order to trick the trained model into failing to correctly filter emails containing spam.

- **model extraction**, obtaining information by extracting information about the particular model used, such as its architecture or parameters;
- **property inference**, accessing global information on the distribution of training data by interacting with the model.

To mitigate reconstruction attacks, it has been proposed that privacy-enhancing techniques be used, such as *differential privacy*, which – through appropriate data manipulation – sets a limit on how much an attacker, with access to the algorithm's outputs, can infer about each individual record in the dataset.

Possible mitigation strategies also include limitations on the number of user queries to the model and the detection of suspicious queries to the model.

11.1.4. Abuse attacks

This category of attacks aims to alter the behaviour of a generative AI system in order to adapt it to the attacker's own ends, such as committing fraud, distributing malware and manipulating information.

Possible mitigation strategies involve the use of methods such as reinforced learning with human feedback, input filtering or the detection of abnormal output values (outliers).

11.2. Cyber risk management

As noted in Section 11.1., AI systems are systems characterised by specific risks whose management is an essential element for the responsible development and use of AI.

Article 15 of the AI Act stipulates that high-risk AI systems must be designed and developed in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and that they perform consistently in those respects throughout their life cycle.

An effective structuring of a risk management process for an AI system must necessarily consider the distinctive characteristics of these systems. For this purpose, reference can be made to specialised frameworks and standards, such as the AI Risk Management Framework (AI RMF)[65] developed by the *National Institute of Standards and Technology (NIST)*.

[65] National Institute of Standards and Technology (NIST), Artificial Intelligence Risk Management Framework (AI RMF 1.0), 2023, https://doi.org/10.6028/NIST.AI.100-1.

The framework is designed to assist organisations and individuals, defined as [66]*AI Actors,* and provides a structured approach to identify, assess and mitigate the risks associated with AI systems. The aim is to minimise potential negative impacts and maximise positive impacts so as to have trustworthy AI systems.

Potential negative impacts arising from the use of AI systems are categorised according to the type of actor involved:

- **impacts on individuals**, such as impacts on civil liberties, physical/psychological safety or an individual's economic circumstances;
- **impacts on organisations**, such as impacts on business operations, reputation or resulting from a breach of an organisation;
- **impacts on the ecosystem,** such as impacts on interconnected and interdependent elements and resources, the financial system, the supply chain or natural resources and the environment.

The 'Core' of the framework identifies the following four functions (in turn divided into categories and subcategories) to support organisations in managing the risks posed by AI systems.

- **GOVERN:** cultivate a culture of risk management within organisations designing, developing, acquiring and deploying AI systems.
- **MAP**: establish the context to frame risks related to an AI system; identify the risks and their associated risk factors.
- **MEASURE**: analyse, assess, benchmark and monitor the risk and related impacts. It uses the information acquired from the previous function and provides guidance to the next one.
- **MANAGE:** set priorities and take action on the identified risks. Risk treatment comprises plans to respond to, recover from, and communicate about incidents or events.

The GOVERN function is cross-cutting that is infused throughout AI risk management. The MAP, MEASURE and MANAGE functions comprise components of the GOVERN function (in particular those related to compliance or assessment) and can be used in specific contexts and at certain stages of the life cycle of AI systems.

[66] The Organisation for Economic Co-operation and Development (OECD) defines *AI Actors* as those who play an active role throughout the AI system life cycle which can include organisations and individuals that deploy or operate AI.

To support organisations in using the Framework, NIST has developed a playbook[67] tailored to each subcategory of the four functions.

Risk identification typically begins with the threats to which a system and its assets may be exposed, and which seek to exploit vulnerabilities in those assets.

For this reason, the following sections list the asset and threat categories related to AI systems based on the life cycle model.

11.3. Assets

An AI system consists of a set of *assets*. An asset is defined as anything that has value to an individual or an organisation and which must therefore be protected. In addition to AI assets (such as models and hyperparameters), ICT infrastructure assets (such as communication networks and operating systems) are also considered here.

The assets of an AI system can be categorised as follows, with the corresponding stage of the life cycle of the AI system given in brackets:

- **data**, such as: raw data (data acquisition), assessment data (model tuning), labelled data (data pre-processing), test data (model training);
- **models,** such as: algorithms (model training), hyperparameters (model tuning), training algorithms (model selection);
- **actors**, such as: cloud service providers (data collection, model training, model tuning), data owners (defining objectives, data collection, data exploration), data providers (data collection);
- processes, such as: data acquisition (data collection), data labelling (data pre-processing), data understanding (data exploration and validation);
- tools, such as: communication networks (data collection), databases (data collection), operating systems (model deployment, model maintenance), interfaces (user and management) and APIs (external);
- **artefacts**, such as: data management policies (data collection), model architecture (model selection, model distribution), use cases (business understanding).

It is essential to carry out systemic mapping of the critical assets that make up the AI system. This makes it possible to track and monitor the most relevant resources, such as data, models, infrastructure and operational processes,

[67] NIST AI RMF Playbook https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook.

ensuring that they are adequately protected throughout all stages of the system's life cycle.

11.4. Threats

Each stage of the life cycle is characterised by one or more threats[68]. In accordance with the ENISA document 'Artificial Intelligence Cybersecurity Challenges', the following categories of threats can be identified:

- **nefarious activity/abuse**[69]: intended actions that target ICT systems, infrastructure, and networks by means of malicious acts with the aim to either steal, alter, or destroy a specified target (e.g. data poisoning and backdoors in the model);
- **eavesdropping/interception/hijacking**: actions aiming to listen, interrupt, or seize control of a third party communication without consent (e.g. disclosure of the model and data theft);
- **physical attacks:** actions which aim to destroy, expose, alter, disable, steal or gain unauthorised access to physical assets such as infrastructure, hardware, or interconnection (e.g. model sabotage and DDOS).
- **unintentional damage:** unintentional actions that cause destruction, harm or injury to property or persons and result in a failure or reduction in usefulness (e.g. reduced data accuracy and impaired selection of characteristics);
- **failures or malfunctions:** partial or full insufficient functioning of a hardware or software asset (e.g. data scarcity and degradation of model performance);
- **outages:** unexpected disruptions of service or decrease in quality falling below a required level (e.g. Infrastructure/system outage, telecommunications network outage);
- **disaster:** a sudden accident or a natural catastrophe that causes great damage (e.g. natural disasters and climate change phenomena);
- **legal:** legal actions of third parties (e.g. due to disclosure of personal information and user profiling).

As AI technologies continue to evolve, new vulnerabilities and risks may emerge, often in an unpredictable manner. Therefore, organisations need to equip

[68] The same threat may refer to multiple life cycle stages.

[69] Section 11.1. deals with the various categories of attacks associated with this type of threat.

themselves with processes and tools to promptly identify emerging threats, analyse them and adapt their defence strategies. The monitoring of emerging threats does not only concern traditional attack techniques, but also includes the detection of new types of exploits based on technological advances, regulatory changes or new ways of interacting with the systems.

11.5. Security objectives

This section sets out the security objectives that should guide the public administration's adoption of AI.

The objectives are borrowed from the *Guidelines for secure AI system development*[70] promoted by the *National Cyber Security Centre of the United Kingdom (NCSC)* and which 35 Agencies from 18 countries have endorsed, including the Italian National Cybersecurity Agency[71].

The recommendations provided for achieving the aforementioned objectives should be regarded as basic guidelines, the implications of which must be carefully understood and assessed during implementation.

It is up to each public administration, based on its specific threat exposure and its own risk analysis, to identify and subsequently achieve further security objectives aimed at strengthening the cybersecurity of its AI systems.

1. **Release AI responsibly.** AI models, applications and systems must be released only after they have undergone a security assessment. Users should also be clearly informed of any known limitations, potential failures, the security aspects for which they are responsible and how (and where) their data could be used, consulted or stored (e.g. if it is used to retrain the model or if it is reviewed by PA employees or third parties).
2. **Identify, track, maintain and protect assets.** Processes and tools must be put in place to identify, track, maintain and protect the assets[72] of AI systems, including, for example, specific controls to manage and protect data and content generated by AI systems. Controls must also be implemented to protect the confidentiality, integrity and availability of AI system logs.
3. **Secure the supply chain.** The supply chain security of the various components of AI systems (e.g. data and models, software libraries, external APIs, etc.) must be assessed and monitored throughout the entire life cycle of AI systems. Components must be acquired from verified providers and must be adequately protected and documented. It is essential to analyse the management of data by third parties, making sure

[70] National Cyber Security Centre (NCSC), Guidelines for secure AI system development.
[71] The guidelines are published on the ACN website at https://www.acn.gov.it/portale/linee-guida-ia.
[72] For the list of asset categories, please refer to Section 11.3..

that adequate security measures are implemented, such as encryption and access controls. Furthermore, providers must be required to ensure that their security measures are in line with the security policies adopted and the risk assessment carried out.

4. **Protect the model and data.** AI system models and data must be protected against unauthorised access or tampering. An attacker may be able to reconstruct the functionality of a model or the data on which it was trained, either by accessing it directly (by obtaining the model weights) or indirectly (by querying the model via an application or service). Attackers can tamper with models, data or requests during or after the model training phase, rendering the outputs unreliable. You can protect the model and data from unauthorised access by implementing cybersecurity best practices and controls on the model query interface to detect and prevent attempts to access, modify or exfiltrate information.
5. **Monitor the behaviour of the system and inputs.** The outputs and performance of the AI model and system must be monitored so that sudden and gradual changes in behaviour that affect security, such as potential intrusions and compromises, can be detected and addressed. In line with data protection requirements, including those relating to personal data, inputs to AI systems (such as inference requests and queries) must be monitored and logged to enable compliance checks, audits, analysis and recovery in the event of a breach or misuse.
6. **Develop an incident response plan.** An incident response plan reflecting the different threat scenarios should be defined and implemented. The plan must be reviewed periodically and in response to internal events (such as updates to strategic plans or organisational changes), external events (such as changes to the regulatory and legislative framework) or changes in the exposure to threats and associated risks, which are regularly assessed in line with developments in the system and in research in general.
7. **Train and raise awareness of threats and risks among staff.** The actors responsible for the adoption of AI must understand threats and related mitigations. Data scientists and developers must be trained in secure software development and in safe and responsible AI practices. For

further information on training and skills development, please refer to Chapter 8. of these guidelines.

8. **Secure the ICT infrastructure.** Any ICT infrastructure hosting AI systems must be adequately protected. Vulnerabilities in ICT infrastructure can, in fact, be exploited by an attacker to carry out attacks on the AI system (such as compromising the model or degrading its performance). Therefore, cybersecurity measures appropriate to the associated risks must be implemented across the ICT infrastructure used throughout the entire life cycle of AI systems. For the cybersecurity measures applicable to ICT infrastructure, subject to the applicable legislation, reference may be made to the guidelines for strengthening resilience referred to in Article 8 of Law No 90 of 28 June 2024 issued by the ACN.
9. **Protect identities and access.** Identity and access management (IAM) is essential for protecting AI systems, particularly when handling sensitive data. The main associated risks include unauthorised access, data theft and model manipulations, which can compromise the integrity and confidentiality of information. In order to mitigate these risks and maintain the overall security of AI systems, it is critical to implement centralised controls that enable effective access monitoring, clearly define roles and permissions to ensure appropriate access and use multi-factor authentication (MFA) to add an additional layer of protection.

11.6. Integrated security management of AI systems

In order to effectively pursue the security objectives outlined in the previous section, the security management of AI systems should be integrated into a cycle that includes at least the following steps, as shown in Figure 7.

Figure 7. AI system security management cycle.

1. Sicurezza in tutte le fasi del ciclo di vita dell'AI	1. Security at every stage of the AI life cycle
2. Audit regolari e controlli di conformità	2. Regular audits and compliance checks
3.Revisione e aggiornamento continuo delle misure di sicurezza	3. Continuous review and updating of security measures

1 **Security at every stage of the AI life cycle:** security must be integrated as a central element throughout the life cycle of AI systems, from defining the objective to maintaining the model. This involves ensuring that the security objectives outlined in the previous section are applied throughout the AI life cycle. Best practices include data protection, threat analysis and the implementation of specific countermeasures for each stage, as shown in the table below. The following are the security best practices for each stage of the life cycle of an AI system.

1.1 stage: **planning and design**

security practices: establish criteria for the security and protection of personal data starting as early as this stage in order to ensure the objective is met with respect to regulatory and operational requirements.

1.2 stage: **data collection and processing**

security practices: apply personal data protection techniques, such as anonymisation, pseudonymisation and encryption. Verify the security of external sources and the reliability of data providers. Ensure that data, especially personal data, is processed in accordance with the principles of data minimisation and integrity. Protect access during the

processing of such data. Ensure that data is protected from unauthorised changes during this stage.

1.3 *stage*: **build and/or train models**
security practices: ensure that sensitive information is not exposed and that processed data is protected from unauthorised access. Ensure that the model is sufficiently robust against adversarial attacks (vulnerabilities may vary depending on the model; consider including details in the relevant sections). Ensure that the data used for the training is free of contamination or bias. Check that changes to the hyperparameters do not expose the model to new vulnerabilities. Protect the tuning process from unauthorised access. Verify the security of pre-trained models and their provenance to avoid the insertion of vulnerabilities through learning transfer. Apply protection techniques against specific attacks to transferred models.

1.4 *stage*: **make available for use or deploy**
security practices: implement access control systems for distributing and protecting the model. Use digital certificates to ensure the integrity of the model that is deployed.

1.5 *stage*: **operate and monitor**
security practices: constantly monitor the behaviour of the model to detect any anomalies or compromises. Use logging systems to track changes and accesses to the model. Carry out a performance analysis of the model in controlled environments to ensure that it continues to operate securely and reliably.

1.6 *stage*: **retire and decommission**
security practices: define and adopt procedures for the safe disposal of retired and/or decommissioned components, such as cleaning the model data.

2 **Regular audits and compliance checks:** organisations must conduct internal and external audits to verify compliance with corporate regulations and policies, with a focus on data protection, model security and compliance with international security standards. These audits must include penetration test activities and attack simulations to assess the resilience of the system.

3 Due to the rapidly changing nature of threats, security measures need to be continuously updated. It is necessary to conduct regular risk assessments and

adopt new protection technologies (such as the introduction of new encryption algorithms or the use of advanced protection techniques against *adversarial* attacks).

<<End of document>>