



Group Regulation on the Ethics and Governance Process of Artificial Intelligence Systems

Modena, 20/12/2023

Version for public disclosure

Version 01

1 General aspects

Summary of the main topics:

This document describes the **Ethics and Governance Process of Artificial Intelligence Systems** adopted by the BPER Group to ensure that the Artificial Intelligence (AI) Systems developed internally are efficiently governed.

The governance guidelines that inspire this process are intended to incorporate the highest benchmarks defined by national and international guidelines and pave the way for the incorporation of guidance from forthcoming **supranational standards** (with specific reference to the AI Act - Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence).

In light of the above, this document formalises the **tools, processes** and **organisational model** aimed at pursuing the above-mentioned objectives.

The document deals with the following topics:

- general definition of the process that all of the organisational units of the BPER Group, without exclusion, must follow for the development and management of AI Systems during their entire life cycle (from business requirements definition, through performance monitoring after release into production, to decommissioning), so that they comply with the standards required by the internal and external regulatory framework;
- guidance on the BPER Group's internal regulatory requirements to be complied with to ensure that the output of all AI Systems is in line with the BPER Group's moral and ethical values;
- identification of all parties that need to be involved during the development of AI systems, according to business, governance and IT perspectives.

2 Definitions

- **Artificial intelligence (AI) or AI System** - an IT system installed on any machine that is designed to operate with a variable degree of autonomy and that, based on a set of explicit or implicit objectives, may generate outputs such as predictions, recommendations or decisions, influencing real or virtual environments¹.
- **Generative Artificial Intelligence (GenAI)** – is a specific subset of AI Systems capable of generating new content that resembles its training data (typically from large and complex pre-trained AI models known as “Foundation Models”). For the purpose of this Regulation, unless otherwise specified, the latter is subject to the same requirements as those applying to AI systems.
- **Bias** - systemic difference in treatment of certain objects, people or groups in comparison to others².
- **Fine-tuning** – for the purposes of this Regulation, fine-tuning shall be understood as an approach to train a pre-trained model on a minimum amount of new data compared to the dataset that would be necessary to build a pre-trained model. This activity is typically needed to fine-tune a pre-trained model and improve its performance in a specific functional setting.
- **Key Performance Indicators (KPIs)** - quantitative indicators structured into the four key dimensions of AI development: Value, Business, Ethical and Technical. These KPIs are monitored over time with a predetermined frequency.

¹ The definition of AI is based on the *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, the (EU) AI Act*.

² The definition of bias is based on the standard *ISO/IEC 24027:2021*.

- **KPI Value** - quantitative indicator that measures the economic value generated by the outputs of the AI System.
- **KPI Business** - quantitative indicator that measures the benefit generated by the AI System for the requesting business structure.
- **KPI Technical** - quantitative indicator that measures the technical performance (in mathematical terms) of the AI System.
- **KPI Ethical** - quantitative indicator that evaluates the ethical performance of the AI System according to the internal and external regulatory frameworks in force and any ethical variables deemed “sensitive” (e.g. gender, age, geography).
- **Pre-trained model** – for the purposes of this Regulation, a pre-trained model shall be understood as an AI system made available by a third party (i.e. Open Source or external providers), which is not operated by learning on datasets but only requires the development of additional applications to feed inputs and obtain outputs.
- **Inference pipeline in an AI system** – a set of instructions in a source code which is fed by the AI system input data and is run to generate an output. The pipeline corresponds to a run (execution) of the AI system with no dependencies on the Monitoring and Training/re-training pipelines.
- **Monitoring Pipeline in an AI system** – a set of instructions in a source code which, if properly fed, generate KPIs to monitor the AI system. The Monitoring pipeline is run with no dependencies on the Monitoring and Training/re-training pipelines according to business needs.
- **Training/re-training pipeline in an AI system** - a set of instructions in a source code which is fed by a dataset and allows the operator to train an AI system that will be integrated into the Inference pipeline. The training/re-training pipeline may be run again to re-train the AI system and revert it back to being efficient when it has become obsolete. This pipeline is run with no dependencies on the Monitoring and Inference pipelines, according to business needs.
- **Human oversight protocol** – for the purposes of this Regulation, human oversight refers to the type of human intervention any AI System is to be subject to. Three possible protocols for human oversight have been identified:
 - **Human-in-Command:** HIC is the most stringent approach, in which the human being has full capability to consider, modify or disregard each and every output generated by the AI system.
 - **Human-in-the-loop:** HITL is a protocol in which every single output generated by the AI system is validated by the human being. Human intervention is thus required for the business process to proceed. In contrast to the Human-in-command protocol, the operator under the HITL approach can either accept or reject (and therefore modify) the outputs generated by the AI System. However, the operator cannot disregard the output by acting in full autonomy, as human intervention and the AI System are integrated in the business process.
 - **Human-on-the-loop:** HOTL is a protocol that does not require each and every output to be validated by the human being. The AI System is fully automated and independent of human intervention in output generation. At high level, this system is monitored by the human being via synthetic KPIs. In case of emergency, the human being can interrupt the execution of the AI System.
- **Re-training** - recalibration of the AI System parameters on the basis of more recent data to maintain an optimal level of performance.

3 Rules and Methodologies

BPER’s Ethics and Governance Process of AI Systems defines the organisational model to be followed for developing, engineering and monitoring AI Systems. As part of this process, the CDO coordinates the activities of the communities of Data Scientists and Citizen Data Scientists, who are part of the Data Science Labs of the different organisational structures of the Group, by defining the guidelines, methodological standards, tools and level of control required for the individual use cases.

This process takes into close consideration all the aspects associated with the algorithm ethics, with a view to

promoting a systematic, integrated and scalable adoption of Ethical AI in the BPER Group, by incorporating and implementing the guidelines of main governmental and non-governmental institutions, including the European Union, UNESCO, the OECD (see para. 3.1.).

The process is also strongly inspired by regulatory developments in AI, with specific reference to the soon-to-be issued AI Act - Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence.

It should also be noted that this Regulation is reassessed at least once a year by the Process Owner to ensure timely alignment of its content with the process, if necessary.

3.1 AI Governance framework

The AI Governance framework considers a multitude of elements for the classification of AI Systems according to risk-based logics, in line with the provisions of the EU AI ACT. Firstly, the AI Systems are identified on the basis of the definition provided by the European Regulator in the EU AI ACT, which is recalled, as is, in the "AI System" definition provided in this Regulation. The Governance Framework proposed herein and, more generally, this Regulation shall be understood as referring to the AI Systems included in this definition.

3.1.1 Risk-based classification of AI Systems

The first classification approach adopted by the BPER Group is based on the EU AI ACT, which enables to associate any AI System with one of the following levels of Risk:

1. Prohibited AI System
2. High-risk AI System
3. Non-high-risk AI System, with Transparency obligations
4. Minimal risk AI System

This Regulation is intended, in the first instance, to incorporate this classification, based on the rationale set forth by the European Regulator in the EU AI ACT to ensure that the Group's AI Systems meet every requirement established by the Law.

3.2 AI Systems Monitoring

This regulation requires that a dedicated Monitoring System be in place for every AI System. Every AI System is more vulnerable to technical obsolescence than traditional software. Obsolescence occurs when real data significantly differs from the data used for training the model, which no longer has the ability to correctly generalise the most recent behaviours. In this case, the model needs to be updated by either re-training or, in the most important cases, changing the way the model operates. The Monitoring System is also functional to ensure AI risk control, along with the verification and ongoing observance of ethical and compliance requirements.

To promptly intercept any instance of obsolescence, it is essential to have an effective Monitoring System in place that can measure the performance of the AI System from a multidimensional perspective. In this regard, the Group has identified four families of KPIs that allow for a synthetic analysis of the different performance features of the AI System:

- Value: in Euro (€), these KPIs measure the economic upside generated by the AI System over time.
- Business: these KPIs measure the impact of the AI System on the business process it is integrated into.
- Technical: these KPIs measure the mathematical performance of the AI System based on its type and technical features.
- Ethical: these KPIs measure the impact of the AI System on natural persons and groups of people to exclude potential discrimination and inequality in forecasts that may be based on personal attributes.

3.3 Ethical Framework for AI Governance

AI Governance activities are based on an Ethical Framework inspired by the following sources:

- the BPER Group values
- the BPER Banca Code of Ethics
- the BPER Group’s ESG (Environmental, Social and Governance) Policy
- the BPER Group’s Personal Data Protection Policy
- the OECD Guidelines- (“Recommendation of the Council on Artificial Intelligence”)
- the G20 Guidelines (“G20 AI Principles”)
- the UNESCO Guidelines (“Recommendation on the Ethics of Artificial Intelligence”)
- the “Universal Guidelines for Artificial Intelligence”
- the Guidelines of the Council of Europe (“Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law”)
- the Guidelines of the European Commission (“Ethics Guidelines for Trustworthy AI”)

The Ethical Framework for AI Governance (Figure 1) is based on the principles of these sources of inspiration and thereby defines the four pillars of AI Governance:

- **Accountability:** the allocation of responsibility in the context of AI Systems
- **Fairness:** the measure of equity, fairness and objectivity of AI Systems aimed at minimising bias or discrimination against individuals
- **Transparency:** the measure of the explainability of AI System outputs for users and data subjects
- **Effectiveness:** the measure of the effectiveness of AI Systems and the level of their integration into corporate processes

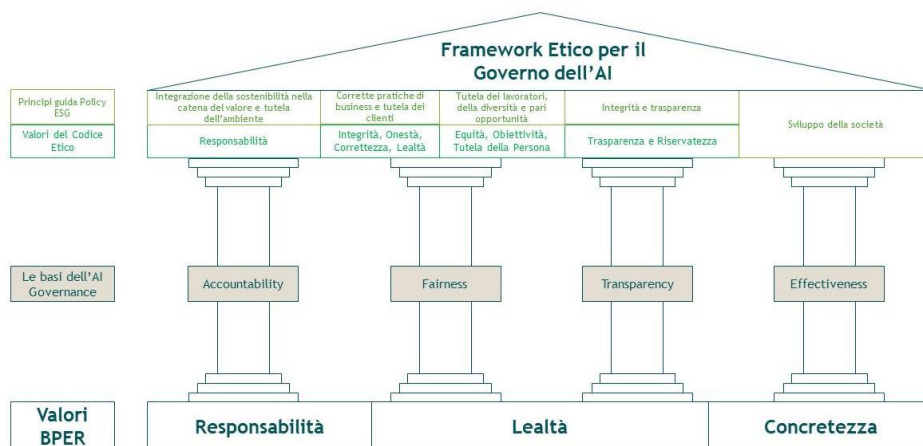


Figure 1 - Ethical Framework for AI Governance

4 Structure of the Ethics and Governance Process of Artificial Intelligence Systems

The Ethics and Governance Process of AI Systems is structured into the following two sub-processes:

- Development of an Artificial Intelligence System
- Monitoring of the Artificial Intelligence Systems

4.1 Sub-process: Development of an Artificial Intelligence System

The “Development of an Artificial Intelligence System” sub-process formalises the roles and responsibilities for project activities involving the design, development and engineering of an AI System; it includes the following phases:

- Use Case Design and Planning
- Model Development
- Data Integration
- Environment Setup
- Go to Production
- Data Visualisation

4.1.1 Phase: Use Case Design and Planning

This phase defines the roles and responsibilities involved in formalising and planning project activities, verifying the AI System technical feasibility and designing AI control mechanisms.

4.1.2 Phase: Model Development

This phase defines the roles and responsibilities involved in the development and approval of a new AI System, which must comply with the requirements set by the Model Owner and BPER’s technical and methodological standards for the fair, efficient and safe development of AI Systems.

4.1.3 Phase: Data Integration

This phase provides for the integration of data from inputs and outputs of the AI System into the Corporate Information System with the methods already identified by other IT processes for software development and release into production.

4.1.4 Phase: Environment Setup

This phase defines the roles and responsibilities involved in the correct setup of the environment for the development, validation, testing and production of AI Systems.

4.1.5 Phase: Go to Production

This phase defines the roles and responsibilities involved in the engineering and release of the AI System into production.

4.1.6 Phase: Data Visualisation

This phase defines the roles and responsibilities involved in implementing an AI monitoring system, consisting of dashboards featuring metrics and graphs that help visualise the AI System’s outputs and KPIs.

4.2 Sub-process: AI Systems Monitoring

The “AI Systems Monitoring” sub-process identifies the activities carried out by the BPER Group to monitor the AI Systems, by measuring their performance on an ongoing basis and enabling all activities for the management of the risk model associated with each AI System. This sub-process is based on the principles set forth by the technical standard “ISO/IEC 38507 - Information Technology - Governance of IT - Governance implications of the use of artificial intelligence by organizations”, which identifies the main elements for the effective governance of AI Systems in production; it comprises the following phases:

- KPI Monitoring
- Application Maintenance and Issue Remediation
- Model Re-train

4.2.1 Phase: KPI Monitoring

This phase defines the roles and responsibilities involved in the continuous monitoring of the AI System’s KPIs during its entire life cycle.

4.2.2 Phase: Application Maintenance and Issue Remediation

This phase defines the roles and responsibilities involved in setting up the AI Systems’ Application Maintenance services and Issue Remediation activities.

4.2.3 Phase: Model Re-train

This phase defines the roles and responsibilities involved in retraining the AI System.

4.3 Regulatory framework

External regulations:

- Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (AI Act), Brussels, 21 April 2021 – COM (2021) 206.
- Regulation (EU) No 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.

Technical standards of reference:

- ISO/IEC 23053:2022 - Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).
- ISO/IEC 22989:2022 - Information technology — Artificial intelligence — Artificial intelligence concepts and terminology.
- ISO/IEC 38507:2022 - Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations.
- ISO/IEC 24027:2021 - Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making.
- ISO/IEC 23894:2023 - Information technology — Artificial intelligence — Guidance on risk management.