Regolamento di Gruppo
del Processo di Etica e Governo dei Sistemi di Intelligenza Artificiale

Modena, *20/12/2023*

Versione per la divulgazione esterna al Gruppo Bancario

Versione 01

1 Aspetti generali

Sintesi delle principali tematiche trattate:

Il presente documento descrive il **Processo di Etica e Governo dei Sistemi di Intelligenza Artificiale** adottato dal Gruppo BPER al fine di garantire che i Sistemi di Artificial Intelligence (AI) sviluppati internamente siano efficacemente governati.

Le linee guida di governo a cui si ispira tale processo mirano a recepire i più alti benchmark definiti dalle linee guida nazionali ed internazionali, nonché a predisporsi nell'ottica di poter recepire le indicazioni delle **regolamentazioni sovra-nazionali** di prossima emanazione (con particolare riferimento all' AI Act - Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale).

Tutto ciò considerato, questo documento formalizza gli **strumenti,** i **processi** e il **modello organizzativo** atti a perseguire le finalità su indicate.

Nel dettaglio il documento tratta le seguenti tematiche:

- la definizione generale del processo a cui tutte le unità organizzative del Gruppo BPER devono, senza esclusione, attenersi per lo sviluppo di Sistemi di AI, e per la gestione di questi ultimi durante l'intera durata del loro ciclo di vita (dalla definizione dei requisiti di business, al monitoraggio a valle del rilascio in produzione, alla dismissione), in modo tale che essi risultino in linea con la regolamentazione imposta dalle normative interne ed esterne;
- l'indicazione dei vincoli normativi interni al Gruppo BPER da rispettare al fine di garantire che il risultato prodotto da ogni Sistema di AI sia affine ai valori etici e morali del Gruppo BPER;
- l'identificazione di tutti gli attori da coinvolgere durante lo sviluppo dei Sistemi di AI, secondo le prospettive di business, di governo e IT.

2 Definizioni

- **Artificial Intelligence (AI) o Sistema di AI** – un sistema informatico installato su una qualsiasi macchina che sia stato progettato per operare con un variabile livello di autonomia e che possa, sulla base di obiettivi espliciti od impliciti, generare output quali predizioni, raccomandazioni o decisioni che influenzino un ambiente fisico oppure virtuali[1].
- **Artificial Intelligence Generativa (GenAI)** – è uno specifico sottoinsieme dei Sistemi di AI, in grado di generare nuovi contenuti simili a quelli osservati in fase di addestramento (effettuato, tipicamente, a partire da modelli di AI pre-addestrati molto grandi e complessi che prendono il nome di "Foundation Model"). Ai fini del presente regolamento, e se non diversamente specificato, ereditano le medesime prescrizioni che si applicano ai Sistemi di AI.
- **Bias** – differenza sistematica nel trattamento di determinati oggetti, persone o gruppi rispetto ad altri[2].
- **Fine-tuning** – ai fini del presente regolamento, si intende una attività di addestramento di un Modello Pre-addestrato con una quantità di dati aggiuntivi minimale rispetto a quella necessaria per la costruzione del Modello Pre-addestrato. Tale attività occorre, generalmente, per affinare il comportamento di un Modello Pre-addestrato e migliorare le sue performance all'interno di un contesto funzionale specifico.
- **Key Performance Indicator (KPI)** – indicatori a carattere quantitativo articolati nelle quattro dimensioni chiave per lo sviluppo di un Sistema di AI: Value, Business, Ethical Technical. Tali KPI sono monitorati nel tempo secondo periodicità predefinita.

[1] La definizione di AI è ispirata alla *proposta di regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'unione, c.d. (EU) AI Act*.

[2] La definizione di bias è ispirata allo *standard ISO/IEC 24027:2021.*

- **KPI Value** – indicatore quantitativo che misura il valore economico generato dagli output del Sistema di AI.
- **KPI Business** – indicatore quantitativo che misura il beneficio generato dal Sistema di AI nei confronti della struttura di business richiedente.
- **KPI Technical** – indicatore quantitativo che misura la performance tecnica (in termini matematici) del Sistema di AI.
- **KPI Ethical** – indicatore quantitativo che valuta la performance etica del Sistema di AI in base alle normative interne ed esterne vigenti di riferimento e ad eventuali variabili etiche ritenute "sensibili" (ad es. sesso, età, provenienza geografica).
- **Modello Pre-addestrato** – si intende, ai fini del presente regolamento, un Sistema di AI reso disponibile da terze parti (ad es. Open Source o fornitori esterni) e per il cui funzionamento non sia necessario alcun tipo di apprendimento su dati, ma sia necessario unicamente costruire le integrazioni applicative di alimentazione degli input e raccolta degli output.
- **Pipeline di Inference di un Sistema di AI** – insieme di istruzioni in codice sorgente che, alimentate dai dati di Input del Sistema di AI, lo eseguono generandone gli output. Tale pipeline corrisponde all'attività di run (esecuzione) del Sistema di AI che avviene senza dipendenze dalle pipeline di Monitoring e di Training / re-training.
- **Pipeline di Monitoring di un Sistema di AI** – insieme di istruzioni in codice sorgente che, opportunamente alimentate, permettono di generare i KPI di monitoraggio del Sistema di AI. Tale pipeline è eseguita senza dipendenze dalle pipeline di Monitoring e di Training / re-training, in linea con le esigenze di business.
- **Pipeline di Training / re-training di un Sistema di AI** - insieme di istruzioni in codice sorgente che, alimentate da un set di dati permette di addestrare il Sistema di AI che sarà integrato nella pipeline di Inference. Tale pipeline può essere ri-eseguita per effettuare il re-train del Sistema di AI e riportarlo in efficienza in caso di obsolescenza tecnica. Tale pipeline è eseguita senza dipendenze dalle pipeline di Monitoring e di Inference, in linea con le esigenze di business.
- **Protocollo di Supervisione Umana** – ai fini del presente regolamento si intende la tipologia di supervisione umana a cui ogni Sistema di AI deve essere sottoposto. Sono individuati tre possibili protocolli di supervisione umana:
    - **Human-in-Command**: è il protocollo di supervisione più stringente in quanto l'essere umano ha piena facoltà di considerare, modificare o ignorare ogni singolo output generato dal Sistema di AI.
    - **Human-in-the-loop**: è un protocollo che prevede che ogni singolo output generato dal Sistema di AI sia validato dall'essere umano. L'intervento umano è pertanto richiesto affinché il processo di business possa proseguire. Diversamente dal protocollo Human-in-command, l'operatore può accettare o rifiutare (e quindi modificare) gli output generati dal Sistema di AI, ma non può ignorarne l'output agendo in completa autonomia, in quanto sia l'intervento umano, sia il Sistema di AI sono integrati fra loro nel processo di business.
    - **Human-on-the-loop**: è un protocollo che non prevede la validazione di ogni singolo output da parte dell'essere umano. Il Sistema di AI è totalmente automatizzato ed indipendente dall'intervento umano nella generazione degli output. Tale sistema è monitorato, ad alto livello, dall'essere umano mediante KPI di performance sintetici. In casi di emergenza, l'essere umano può interrompere l'esecuzione del Sistema di AI.
- **Re-training** – ricalibrazione dei parametri del Sistema di AI su dati più recenti, al fine di mantenere un livello ottimale di performance del Sistema di AI.

3 Regole e Metodologie

Il processo di Etica e Governo dei Sistemi di AI di BPER definisce il modello organizzativo da osservare per lo sviluppo, l'ingegnerizzazione e il monitoraggio dei Sistemi di AI. Nell'ambito di questo processo, il CDO coordina le attività delle community di Data Scientist e Citizen Data Scientist appartenenti ai Data Science Lab delle varie strutture organizzative di Gruppo definendo le linee di indirizzo, gli standard metodologici, gli strumenti, e il livello di presidio che occorre adottare sui singoli use case.

Tale processo tiene in forte considerazione tutti gli aspetti legati all'etica degli algoritmi, nell'ottica di

promuovere una adozione sistematica, integrata e scalare dell'Ethical AI nel Gruppo BPER, andando a recepire ed implementare le linee guida di indirizzo delle principali istituzioni governative e non governative fra cui l'Unione Europea, l'UNESCO, l'OECD (si veda il par. 3.1.).

Il processo è, inoltre, fortemente ispirato alle evoluzioni normative in materia di intelligenza artificiale, con particolare riferimento all'AI Act - Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale, di prossima emanazione.

Si precisa, inoltre, che il presente Regolamento viene rivalutato con cadenza almeno annuale da parte del Process Owner, al fine di garantire in caso di necessità un tempestivo allineamento dei contenuti con il processo rappresentato.

3.1 Framework di Governo dell'AI

Il framework di Governo dell'AI prende in esame una moltitudine di elementi per la classificazione dei Sistemi di AI secondo logiche risk-based, in linea con quanto indicato dall'(EU) AI ACT. In primo luogo, vengono identificati i Sistemi di AI sulla base della definizione fornita dal Regolatore Europeo nell'(EU) AI ACT e che viene ripresa, tal quale, nella definizione di "Sistema di AI" fornita nel presente Regolamento. Il Framework di Governo qui proposto e, più in generale, questo Regolamento è da intendersi riferito ai Sistemi di AI che rientrino in tale definizione.

3.1.1 Classificazione dei Sistemi di AI sulla base del Rischio

La prima logica di classificazione che il Gruppo BPER ha adottato è individuata sulla base dell'(EU) AI ACT che permette di associare ad un qualsiasi Sistema di AI uno dei seguenti livelli di Rischio:

1. Sistema di AI Proibito
2. Sistema di AI ad Alto Rischio
3. Sistema di AI non ad Alto Rischio, con obblighi di Trasparenza
4. Sistema di AI a rischio irrilevante

Il presente Regolamento mira a recepire, in prima istanza, questo tipo di classificazione, sulla base dei razionali che il Regolatore Europeo sancisce all'interno dell'(EU) AI ACT per assicurare che ogni requisito indicato dalla Legge possa essere attuato sui Sistemi di AI del Gruppo.

3.2 Sistema di Monitoraggio dei Sistemi di AI

Il presente regolamento stabilisce, per ogni Sistema di AI, di disporre un Sistema di Monitoraggio dedicato. Ogni Sistema di AI è, infatti, più suscettibile di obsolescenza tecnica rispetto al software tradizionale. L'obsolescenza avviene nel momento in cui i dati reali differiscano significativamente rispetto ai dati utilizzati per l'addestramento del modello che non ha più capacità di generalizzare correttamente i comportamenti più recenti. In tal caso, si rende necessario l'aggiornamento del modello attraverso il re-train del modello o, nei casi più importanti, attraverso una modifica evolutiva del modello. Inoltre, il Sistema di monitoraggio è funzionale a garantire il presidio dei rischi del Sistema AI e la verifica ed il rispetto nel continuo dei requisiti etici e di conformità.
Per intercettare tempestivamente le casistiche di obsolescenza è indispensabile predisporre un Sistema di Monitoraggio efficace e capace di misurare le performance del Sistema di AI da una prospettiva multidimensionale. A questo proposito, il Gruppo ha identificato quattro famiglie di KPI che permettano una analisi sintetica delle diverse caratteristiche di performance del Sistema di AI:

- KPI Value, misurati in euro (€) permettono di misurare il beneficio economico generato, nel tempo, dal Sistema di AI.
- KPI Business, misurano l'impatto del Sistema di AI sul processo di business a cui esso è integrato.
- KPI Technical, misurano la performance matematica del Sistema di AI sulla base della sua tipologia e delle sue caratteristiche tecniche.
- KPI Ethical, misurano l'impatto del Sistema di AI sulle persone fisiche e sui gruppi di persone per escludere possibili discriminazioni e disparità nelle previsioni che possano essere basate su attributi personali.

3.3 Framework Etico per il Governo dell'AI

Le attività di Governo dell'AI si basano su un Framework Etico avente le seguenti fonti ispiratrici:

- I valori del Gruppo BPER
- Il Codice Etico di BPER Banca
- La Policy in materia di ESG (Environmental, Social and Governance) del Gruppo BPER
- La Policy in materia di Protezione dei Dati Personali del Gruppo BPER
- Le Linee Guida OECD ("Recommendation of the Council on Artificial Intelligence")
- Le Linee Guida G20 ("G20 AI Principles")
- Le Linee Guida UNESCO ("Recommendation on the Ethics of Artificial Intelligence")
- Le "Universal Guidelines for Artificial Intelligence"
- Le Linee Guida del consiglio d'Europa ("Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law")
- Le Linee Guida della Commissione Europea ("Ethics Guidelines for Trustworthy AI")

Il Framework Etico per il Governo dell'AI (Figura 1) si innesta ai principi delle fonti ispiratrici, andando a definire le quattro basi dell'AI Governance:

- **Accountability**: l'assegnazione della responsabilità nell'ambito dei Sistemi di AI
- **Fairness**: la misura dell'equità, correttezza ed obiettività dei Sistemi di AI orientata alla minimizzazione dei fenomeni di distorsione o discriminazione degli individui
- **Transparency**: la misura della spiegabilità degli output dei Sistemi di AI nei confronti degli utilizzatori e degli interessati
- **Effectiveness**: la misura dell'efficacia dei Sistemi di AI e del grado di integrazione con i processi aziendali

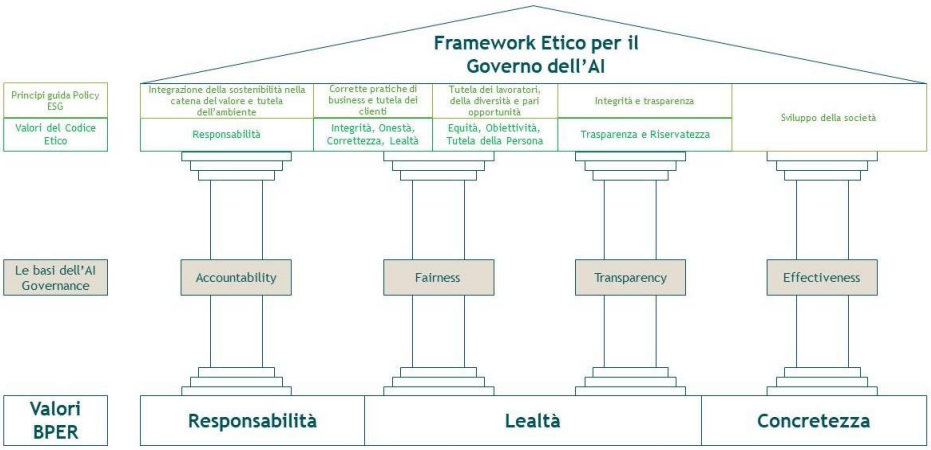

Figura 1 – Framework Etico per il Governo dell'AI

4 Articolazione del processo di Etica e Governo dei Sistemi di Intelligenza Artificiale

Il processo di Etica e Governo dei Sistemi di Intelligenza Artificiale si articola nei seguenti due sotto-processi:

- Sviluppo di un Sistema di Intelligenza Artificiale
- Monitoraggio dei Sistemi di Intelligenza Artificiale

4.1 Sottoprocesso: Sviluppo di un Sistema di intelligenza artificiale

Il sottoprocesso di Sviluppo di un Sistema di Intelligenza Artificiale formalizza i ruoli e le responsabilità per le attività di progetto per l'ideazione, lo sviluppo e l'ingegnerizzazione di un Sistema di AI; si articola nelle seguenti fasi:

- Use Case Design and Planning
- Model Development
- Data Integration
- Environment Setup
- Go to production
- Data Visualization

4.1.1 Fase: Use Case Design and Planning

Questa fase definisce i ruoli e le responsabilità nel formalizzare e pianificare le attività di progetto, verificare la fattibilità tecnica del sistema di AI e progettarne i meccanismi di controllo.

4.1.2 Fase: Model Development

Questa fase definisce i ruoli e le responsabilità nello sviluppo ed approvazione del nuovo Sistema di AI, che deve soddisfare i requisiti stabiliti dal Model Owner e gli standard tecnici e metodologici di BPER per lo sviluppo corretto, efficace e sicuro dei Sistemi di AI.

4.1.3 Fase: Data Integration

Questa fase prevede l'integrazione dei dati di input ed output del Sistema di AI all'interno del Sistema Informativo aziendale con le modalità già individuate da altri processi IT relativi allo sviluppo e alla messa in produzione del software.

4.1.4 Fase: Environment Setup

Questa fase definisce i ruoli e le responsabilità relativamente alla corretta predisposizione degli ambienti di sviluppo, convalida, test e produzione del Sistema di AI.

4.1.5 Fase: Go To Production

Questa fase definisce i ruoli e le responsabilità relativamente all'ingegnerizzazione e al rilascio del Sistema di AI nell'ambiente di produzione.

4.1.6 Fase: Data Visualization

Questa fase definisce i ruoli e le responsabilità nella realizzazione del sistema di monitoraggio dei Sistemi di AI costituito da dashboard contenenti metriche e grafici, al fine di visualizzare gli output ed i KPI del Sistema di AI.

4.2 Sottoprocesso: Monitoraggio dei Sistemi di Intelligenza Artificiale

Il sotto-processo di "Monitoraggio dei Sistemi di Intelligenza Artificiale" identifica le attività svolte dal Gruppo BPER finalizzate al monitoraggio dei Sistemi di AI, misurandone costantemente le performance e abilitando tutte le attività di gestione del rischio modello associato ad ogni sistema di AI. Tale sotto-processo è ispirato ai principi stabiliti dalla norma tecnica "ISO / IEC 38507 - Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations" che identifica gli elementi principali per un governo efficace dei Sistemi di AI in produzione; esso si articola nelle seguenti fasi:

- KPI Monitoring
- Application Maintenance and Issue Remediation
- Model re-Train

4.2.1 Fase: KPI Monitoring

Questa fase definisce i ruoli e le responsabilità relativamente alle attività di monitoraggio continuo dei KPI del Sistema di AI per tutto il suo ciclo di vita.

4.2.2 Fase: Application Maintenance and Issue Remediation

Questa fase definisce i ruoli e le responsabilità nell'ambito della predisposizione dei servizi di Application Maintenance e delle attività di Remediation delle Issue dei Sistemi di AI.

4.2.3 Fase: Model re-Train

Questa fase definisce i ruoli e le responsabilità nell'ambito dell'esecuzione del re-Train (o riaddestramento) di un Sistema di AI.

4.3 Contesto normativo di riferimento

Normativa esterna:

- Bozza del Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (AI Act.), Bruxelles, 21 Aprile 2021 - COM(2021) 206.
- Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE.

Norme tecniche di riferimento:

- 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.