POSITION PAPER

Centre for Law and Ethics of Innovation, Technology and Artificial Intelligence (LEITAI) – Università San Raffaele Roma, Pegaso e Mercatorum

Osservazioni sulla Bozza di Linee Guida per lo sviluppo di sistemi di Intelligenza Artificiale nella pubblica amministrazione

Capitolo 1 - Ambito di applicazione

1. Perimetro soggettivo e principio di proporzionalità

Pur apprezzando l'ampiezza dell'ambito soggettivo e oggettivo, si rileva come l'attuale impostazione rischi di risultare eccessivamente generalista. L'applicazione uniforme a tutte le pubbliche amministrazioni rischia di non riflettere le asimmetrie esistenti in termini di maturità digitale, capacità organizzativa e risorse umane, finanziarie e tecnologiche.

- Criticità: la mancata distinzione tra amministrazioni centrali e locali o tra enti con diversi livelli di digitalizzazione, può generare una duplice distorsione: requisiti insostenibili per le realtà meno strutturate e indicazioni sottodimensionate per quelle più avanzate.
- Proposta: si suggerisce l'introduzione di una classificazione delle PA basata su livelli di maturità (base, intermedio, avanzato). Tale approccio consentirebbe di modulare obblighi e requisiti secondo il principio di proporzionalità, garantendo l'efficacia concreta delle Linee Guida in funzione della reale capacità di esecuzione del singolo ente.

2. Allineamento all'AI Act e prevenzione dell'*overregulation*

L'attuale definizione del perimetro oggettivo appare non perfettamente allineata alla tassonomia del Regolamento (UE) 2024/1689 (AI Act). Il riferimento generico a "tutte le componenti che impiegano tecnologie di IA" rischia di estendere l'applicazione a sistemi algoritmici tradizionali, determinando una superflua *overregulation*.

- Proposta di integrazione (nel §1): È necessario inserire un rinvio esplicito alla definizione di sistema di IA di cui all'art. 3 dell'AI Act.
- Obbligo di classificazione: si propone di prevedere l'obbligo per le PPAA di effettuare una classificazione preventiva di ogni sistema adottato secondo la tassonomia europea, documentandola.

3. Diritti fondamentali come architrave della governance

Si osserva come le Linee Guida, pur trattando temi cruciali quali trasparenza, qualità del dato e sicurezza, omettano un collegamento esplicito con la tutela dei diritti fondamentali.

- Criticità: Tale nesso rappresenta il pilastro dell'impianto regolatorio europeo. Senza un ancoraggio ai diritti fondamentali, risultano privi di giustificazione giuridica essenziale i requisiti relativi alla mitigazione dei *bias*, alla supervisione umana e agli audit etici.
- Proposta: Integrare il testo evidenziando come la conformità tecnica (dataset non discriminatori, *human oversight*) sia strumentale alla prevenzione di impatti negativi sui diritti dei cittadini. Tale prospettiva è indispensabile per legittimare i processi di monitoraggio e la valutazione dell'equità dei sistemi IA nella sfera pubblica.

Capitolo 2 - Come sviluppare sistemi di Intelligenza Artificiale

In merito alla sezione del documento "Come sviluppare sistemi di Intelligenza Artificiale" (capitolo 2), si sottopongono le seguenti proposte di integrazione e revisione strategica.

1. Valorizzazione di architetture locali, modulari e *open source*

Si suggerisce di integrare i paragrafi dedicati allo sviluppo e all'architettura logica con un esplicito riferimento al valore strategico delle architetture locali, modulari e basate su Open Source.

- Motivazione: Per la Pubblica Amministrazione, tale modello rappresenta una garanzia di sovranità digitale, trasparenza algoritmica e indipendenza dai singoli fornitori (*vendor lock-in*). L'adozione di standard aperti facilita, inoltre, la portabilità dei sistemi e la neutralità tecnologica, elementi chiave per l'efficienza nel lungo periodo.

2. Definizione di una gerarchia dei principi e prioritizzazione

Dall'analisi dei venti principi individuati nella bozza, emerge una criticità legata alla loro presentazione paritaria, che non distingue tra requisiti cogenti e indicazioni di miglioramento.

- Criticità: L'assenza di una chiara prioritizzazione rende complesso, per le amministrazioni, identificare i pilastri fondamentali su cui concentrare gli sforzi iniziali di adozione con il conseguente rischio di una dispersione di risorse.
- Proposta: Si ritiene opportuno strutturare i principi secondo una scala di applicabilità progressiva, distinguendo tra principi minimi, essenziali e inderogabili per ogni sistema di IA (es. sicurezza, legalità, protezione dati) e quelli raccomandati, ossia obiettivi di maturità organizzativa e tecnologica.

3. Flessibilità e adozione progressiva

L'articolazione proposta consentirebbe di rendere le Linee Guida uno strumento più flessibile e concretamente applicabile anche dalle amministrazioni meno strutturate. Definendo un perimetro minimo di conformità, si facilita un percorso di adozione incrementale che evita di paralizzare i piccoli enti dinanzi a requisiti eccessivamente complessi, pur mantenendo elevati gli standard per i sistemi ad alto impatto.

Capitolo 3 - Architettura logica di riferimento: orchestratore, modelli, dati e tool

In merito alla sezione sull'architettura logica di riferimento, si rilevano alcune criticità legate alla traducibilità operativa delle indicazioni e alla coerenza con il quadro regolatorio nazionale.

1. Necessità di esempi operativi

L'attuale formulazione del modello architetturale, pur solida concettualmente, presenta un livello di astrazione elevato che ne limita l'immediata applicabilità da parte delle amministrazioni.

- Criticità: l'assenza di schemi di riferimento operativi rischia di rendere difficoltosa la transizione dalla teoria alla realizzazione tecnica delle soluzioni.
- Proposta: si suggerisce di integrare il capitolo con architetture di riferimento, differenziate per casi d'uso tipici della PA e fornire indicazioni tecnologiche, anche in forma esemplificativa e non vincolante, per orientare le PA verso soluzioni scalabili e sicure.

2. Armonizzazione normativa

Nel paragrafo 3.4.1 (Dati di Addestramento) si riscontra un disallineamento terminologico e classificatorio tra la bozza e l'impianto regolatorio definito dall'Agenzia per la Cybersicurezza Nazionale (ACN) in materia di Cloud per la PA.

- Criticità: la mancanza di uniformità con il quadro ACN può generare incertezza giuridica e operativa nella gestione e conservazione dei dati utilizzati per l'addestramento dei sistemi IA.
- Proposta di revisione: si raccomanda un'integrazione volta ad armonizzare integralmente il testo con il Regolamento Cloud ACN, adottando in modo puntuale la classificazione dei dati e dei servizi cloud prevista da ACN, garantendo che le Linee Guida AGID agiscano in sinergia e non in sovrapposizione con le norme di cybersicurezza nazionale già vigenti.

Capitolo 4 - Ciclo di vita dei sistemi di Intelligenza Artificiale e azioni per lo sviluppo e il procurement

Sebbene la bozza delinei con chiarezza i passaggi concettuali del ciclo di vita dei sistemi, si ravvisa la necessità di un maggiore orientamento all'esecuzione e alla governance.

1. Operatività e strumenti di supporto all'implementazione

L'attuale descrizione delle fasi del ciclo di vita risulta priva di un corredo strumentale che ne permetta la traduzione in prassi amministrative standardizzate.

- Criticità: la mancanza di attività di dettaglio, strumenti di supporto e indicazione dei *deliverable* attesi per ogni fase rende complessa l'adozione delle Linee Guida nei processi di *procurement* e sviluppo reale.
- Proposta: si suggerisce di integrare il capitolo con *tool* operativi a supporto dei responsabili e dei dirigenti della PA, quali checklist per ogni fase e template di monitoraggio.

2. Integrazione tra ciclo di vita tecnico e governance organizzativa

Si rileva l'opportunità di rendere più organico il rapporto tra l'evoluzione tecnica del sistema di IA e le strutture di governance dell'ente.

- Obiettivo: rafforzare il collegamento tra il ciclo di vita del software e i modelli di governance organizzativa, definendo chiaramente responsabilità, ruoli e flussi informativi tra le diverse funzioni (tecnica, legale, etica) coinvolte nella gestione del sistema.
- Vantaggio: tale integrazione assicura che il sistema non sia gestito come un elemento isolato, ma come parte di una strategia di innovazione istituzionale coerente e presidiata.

Capitolo 5 - Sicurezza cibernetica

In merito alla sezione sulla sicurezza, pur apprezzando la completezza nella tassonomia delle minacce, si rileva la necessità di un approccio più orientato alla prevenzione attiva e alla conformità tecnica.

1. Operatività e strumenti di mitigazione del rischio

L'attuale formulazione offre una panoramica teorica delle minacce, ma risulta priva di indicazioni prescrittive per l'attuazione delle contromisure.

- Criticità: il divario tra l'identificazione della minaccia e la sua mitigazione pratica rende le Linee Guida di difficile applicazione, in particolare per gli enti con limitate competenze specialistiche in ambito *cyber*.
- Proposta di integrazione: si suggerisce di corredare il capitolo con esempi operativi di mitigazione e di definire una baseline minima di sicurezza per i sistemi di IA, modulata in base ai livelli di rischio identificati.

2. Specificità delle minacce IA e superfici di attacco

Si ravvisa l'opportunità di approfondire le vulnerabilità peculiari dei modelli di Intelligenza Artificiale, distinguendole dalle minacce informatiche tradizionali.

- Azione richiesta: esplicitare le superfici di attacco tipiche dei sistemi IA, fornendo indicazioni puntuali per la difesa del perimetro logico.
- Allineamento ai Framework: si raccomanda di rafforzare il collegamento con i framework di sicurezza nazionali e internazionali già consolidati, garantendo un approccio integrato.

3. Finalità: dalla descrizione alla progettazione sicura

L'integrazione di tali elementi tecnici consentirebbe di trasformare la sezione sicurezza in uno strumento funzionale a tre fasi critiche per la PA:

- Progettazione: adozione di modelli di *Security-by-design*.
- Verifica: conduzione di audit e test di resilienza.
- Procurement: definizione di requisiti di sicurezza stringenti e verificabili nei capitolati di gara.

Capitolo 6 - Neutralità hardware, acceleratori e portabilità dei sistemi di IA

In merito ai requisiti di portabilità e indipendenza tecnologica, si evidenzia la necessità di un approccio che concili l'aspirazione alla piena neutralità con i vincoli operativi del mercato ICT.

1. Realismo operativo e gestione del Vendor Lock-in

L'attuale impostazione sulla neutralità hardware appare eccessivamente ambiziosa rispetto alle reali capacità di implementazione delle amministrazioni.

- Criticità: un'adesione assoluta alla neutralità può scontrarsi con l'efficienza offerta da hardware specializzato o soluzioni proprietarie altamente performanti.
- Proposta: si suggerisce di esplicitare i necessari compromessi architetturali e di introdurre livelli differenziati di neutralità. È fondamentale definire i casi in cui un grado controllato di lock-in sia accettabile.

2. Portabilità degli asset di Generative AI

Il testo limita il concetto di portabilità alla sola infrastruttura, trascurando gli elementi che costituiscono il vero valore strategico dei sistemi di IA Generativa.

- Criticità: il rischio di dipendenza tecnologica è massimo laddove non sia garantita la portabilità di prompt, configurazioni e modelli personalizzati sviluppati nel tempo dalla PA.
- Proposta di integrazione: estendere esplicitamente gli obblighi di portabilità a tutti gli asset logici e informativi, prevedendo standard aperti: obblighi contrattuali per l'esportazione dei dati e delle istruzioni (prompt) in formati interoperabili.

3. Salvaguardia della sovranità digitale

Tali integrazioni sono indispensabili per assicurare che la Pubblica Amministrazione mantenga il controllo sui propri processi decisionali algoritmici, permettendo una migrazione fluida verso soluzioni alternative o nazionali qualora mutassero le condizioni di mercato o le esigenze di sicurezza.