

ITALIA

 OPENGOV

RISCHI E OPPORTUNITÀ DELL'IA PER MIGLIORARE EQUITÀ E ACCESSIBILITÀ

Report sugli esiti del percorso di
co-creazione delle Raccomandazioni
della community italiana per il governo
aperto sull'utilizzo dell'Intelligenza Artificiale



UNIONE EUROPEA
Fondo Sociale Europeo
Fondo Europeo di Sviluppo Regionale



Presidenza del Consiglio dei Ministri
Dipartimento per le politiche di coesione
e per il sud



Presidenza del Consiglio dei Ministri
Dipartimento della Funzione Pubblica

pac PROGRAMMA
OPERATIVO
COMPLEMENTARE

pan GOVERNANCE
E CAPACITÀ
ISTITUZIONALE
2014-2020

INDICE

1. Introduzione e contesto	3
2. Il percorso di co-creazione e i temi degli incontri	4
2.1 Sintesi I incontro di co-creazione	6
2.2 Sintesi II incontro di co-creazione	10
2.3 Sintesi III incontro di co-creazione	13
3. Prossimi passi verso la stesura delle Raccomandazioni	16

Il documento è stato elaborato nell'ambito del Sesto Piano di azione nazionale per il Governo Aperto, obiettivo "B": "Accompagnare la diffusione e l'innovazione delle politiche di apertura a tutti i livelli di governo", impegno 5: "Promozione dell'inclusività e dei diritti nell'accesso alle tecnologie e nell'utilizzo dell'Intelligenza Artificiale".

Organizzazioni responsabili per l'attuazione dell'impegno:

PA: Consiglio Nazionale Giovani (CNG), Regione Puglia

OSC: Federmanager, Period Think Tank

Organizzazioni che collaborano all'attuazione dell'impegno:

PA: Agenzia per l'Italia Digitale (AgID), Autorità Nazionale Anticorruzione (ANAC), Ministero della cultura - Istituto centrale per la digitalizzazione del patrimonio culturale - DIGITAL LIBRARY

OSC: Association of Facilitators Chapter Italia (IAF), Associazione della Comunicazione Pubblica e Istituzionale (Compubblica), Mappina, PASocial

Il presente documento è distribuito con licenza [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) international.



1. Introduzione e contesto

La definizione delle “Raccomandazioni della Community italiana per il governo aperto sull’utilizzo dell’Intelligenza Artificiale” rientra tra le attività previste nell’ambito dell’**impegno B5 - Promozione dell’inclusività e dei diritti nell’accesso alle tecnologie e nell’utilizzo dell’Intelligenza Artificiale** del 6° Piano d’Azione Nazionale per il governo aperto (6NAP) 2024-2026.

Questo report illustra i risultati del processo partecipativo realizzato a sostegno della formulazione e co-creazione delle Raccomandazioni, condivise tra istituzioni pubbliche e organizzazioni della società civile appartenenti alla community italiana di OGP. Il percorso di co-creazione aveva come filo conduttore la **difesa dello spazio civico e la promozione dei diritti nello sviluppo e adozione dei sistemi di intelligenza artificiale**. Il team di attuazione dell’impegno B5, sulla base degli esiti raccolti nel processo, elaborerà successivamente un documento da sottoporre agli enti della pubblica amministrazione per una loro adozione. Il processo di co-creazione, articolato in tre incontri, aveva tra le sue finalità quelle di:

1. co-creare Raccomandazioni della Community italiana per il governo aperto sull’utilizzo dell’Intelligenza Artificiale, per migliorare equità e accessibilità;
2. promuovere il confronto pubblico e maggiore consapevolezza su questo tema e sui potenziali *bias* e discriminazioni dell’Intelligenza Artificiale.

Al processo di co-creazione, organizzato a partire da maggio 2025 e realizzato nei mesi di luglio, settembre e ottobre 2025, hanno preso parte: membri delle organizzazioni appartenenti al team dell’impegno B5, attori-chiave individuati attraverso manifestazione di interesse, soggetti della Community OGP-IT che desideravano partecipare e tutti i cittadini e le cittadine interessati/e.

Complessivamente hanno preso parte attiva al processo circa 40 partecipanti per incontro, per un totale di circa 120 partecipanti, in rappresentanza delle proprie organizzazioni di riferimento.

Con il percorso di co-creazione delle Raccomandazioni e i tre incontri, si è chiusa la fase che precede la definizione di una bozza per le Raccomandazioni, da sottoporre alla consultazione consultazione pubblica online.

L’iniziativa promuove un utilizzo dell’Intelligenza Artificiale equo, inclusivo e rispettoso dei diritti fondamentali della persona umana, rafforzando la consapevolezza sui rischi e le opportunità connessi alla diffusione dei sistemi di IA nella pubblica amministrazione. In coerenza con il quadro europeo delineato dall’AI Act e con il percorso normativo nazionale in materia di IA, l’obiettivo è integrare valutazioni di impatto attente ai profili tecnici, etici e

sociali, con particolare riferimento ai bias di genere, intergenerazionali e intersezionali, contribuendo a rendere le politiche pubbliche più efficaci ed eque.

Tra i principali risultati attesi vi sono il rafforzamento delle competenze nella gestione di processi partecipativi su temi tecnologici complessi, la mappatura di pratiche e criticità legate ai *bias* algoritmici e la definizione di Raccomandazioni concrete per promuovere trasparenza, equità e accountability nell'uso dell'IA nella PA. Un messaggio centrale dell'iniziativa è che l'IA non è neutrale: il modo in cui viene progettata, regolata e utilizzata incide direttamente sulla tutela dei diritti e sulla qualità della democrazia.

Le Raccomandazioni saranno rivolte agli organismi nazionali competenti, alle pubbliche amministrazioni centrali e regionali – anche attraverso il coinvolgimento della Regione Puglia – nonché alla più ampia Community OGP IT, con l'obiettivo di rafforzare il dialogo tra società civile e decisori pubblici e di promuovere un modello di governance dell'IA fondato su partecipazione, trasparenza e responsabilità.

2. Il percorso di co-creazione e i temi degli incontri

Il percorso si è sviluppato nell'ottica del governo aperto e con un processo partecipativo e deliberativo promosso dall'Open Government Partnership Italia, con il supporto metodologico dell'organizzazione internazionale People Powered: per circa 6 mesi si è infatti svolto un percorso di mentoring organizzato nell'ambito del DIA - Democracy Innovation Accelerator Program, a cui hanno partecipato alcuni rappresentanti delle organizzazioni facenti parte del Team di attuazione dell'Impegno. L'esito di tale percorso è stato la progettazione del percorso di co-creazione, che ha avuto lo scopo di coinvolgere amministrazioni pubbliche, organizzazioni della società civile, esperti e altri stakeholder, al fine di costruire in modo condiviso principi etici e raccomandazioni operative sull'equità dei dati e degli algoritmi.

Il percorso si è svolto attraverso tre incontri di co-creazione, ciascuno dedicato a una diversa prospettiva (inquadramento del tema, analisi delle criticità e definizione delle raccomandazioni). La metodologia adottata è stata deliberativa e collaborativa, applicando una logica multi-stakeholder, valorizzando la presenza di membri di amministrazioni pubbliche, società civile, mondo della ricerca, attraverso tavoli di lavoro moderati da facilitatori e facilitatrici, per raccogliere i contributi nel rispetto dei principi di partecipazione, trasparenza, inclusione e accountability propri del governo aperto.

Gli incontri hanno avuto come filo conduttore la promozione dei diritti e la difesa dello spazio civico nello sviluppo e nell'adozione dei sistemi di intelligenza artificiale. I temi-chiave su cui sono state sviluppate le domande-guida sono stati: **consapevolezza, equità, governance e partecipazione, appropriatezza tecnologica.**

Questi temi definiscono un perimetro entro il quale formulare le raccomandazioni, che il gruppo di lavoro responsabile dell'impegno B5 sintetizzerà in un documento da inviare ai decisori della pubblica amministrazione che dovranno garantire l'equità dei sistemi di intelligenza artificiale e l'applicazione dei diritti. L'insieme delle tematiche discusse, durante i tre incontri previsti, potrà contribuire a garantire **l'adozione di un'IA più responsabile e attenta ai diritti fondamentali, all'equità, a un'applicazione estensiva dei processi messi in campo.**

Nello specifico:

- 1) Il primo incontro è stato dedicato al tema della **raccolta, preparazione e validazione dei dati che servono ad alimentare e addestrare il sistema IA.** La qualità dei dati determina la qualità delle decisioni prese dall'algoritmo. Eventuali *bias*, pregiudizi o

discriminazioni presenti nei dati iniziali influenzano inevitabilmente l'intero sistema IA.

- 2) Il secondo incontro ha affrontato il tema dell'**apertura del processo di progettazione, sviluppo e adozione degli algoritmi**, includendo scelte tecnologiche, modalità di funzionamento e trasparenza delle logiche decisionali adottate. In particolare, l'attenzione è stata rivolta all'adozione di soluzioni tecnologiche sul mercato e a discutere di come le PA possono esercitare una funzione di committenza vigile e critica verso la scelta dei fornitori, indicando i requisiti rispetto al procurement per garantire il rispetto dei diritti fondamentali all'interno del loro *"conformity assessment"* (valutazione di conformità).
- 3) Il terzo incontro ha riguardato nello specifico i **sistemi di controllo e verifica dei sistemi di IA, in particolare la fase di addestramento**. In questa fase l'algoritmo viene effettivamente addestrato e affinato attraverso dati reali e anche dati sintetici, testato in contesti concreti e poi costantemente monitorato e aggiornato nel tempo. Qui possono emergere chiaramente eventuali errori, discriminazioni o malfunzionamenti del sistema. L'addestramento richiede attenzione costante per garantire che il sistema funzioni in modo equo e affidabile durante tutto il suo ciclo di vita.

I successivi paragrafi illustrano nel dettaglio le sintesi degli esiti degli incontri.

2.1 Sintesi dell'incontro di co-creazione

Il primo incontro del percorso di co-creazione, dedicato al **tema della raccolta, preparazione e validazione dei dati che alimentano e addestrano i sistemi di intelligenza artificiale**, ha coinvolto rappresentanti della pubblica amministrazione, del mondo accademico e della società civile. Il confronto si è concentrato su alcuni nodi centrali: le modalità di raccolta dei dati, la loro qualità e rappresentatività, i processi di validazione e valutazione di conformità e, in particolare, il ruolo della Pubblica Amministrazione nella governance di questa fase iniziale.

È emersa con chiarezza una consapevolezza condivisa: la qualità dei dati determina la qualità delle decisioni assunte dall'algoritmo. Eventuali *bias*, pregiudizi o discriminazioni presenti nei dati di partenza rischiano di propagarsi lungo l'intero ciclo di vita del sistema di IA, influenzandone i risultati e gli impatti.

La discussione, articolata in gruppi di lavoro paralleli, è stata orientata da una domanda guida comune: **Come possiamo assicurare equità nei dati raccolti, in una fase antecedente allo sviluppo dell'algoritmo, per ridurre discriminazioni e stereotipi?**

L'analisi dei report delle cinque sessioni di gruppo evidenzia otto assi ricorrenti sui quali si registra una forte convergenza tra i partecipanti. Un tema rivelatosi centrale, emerso in tutti i gruppi di lavoro, riguarda **la qualità e la rappresentatività dei dati**. Gli elementi condivisi

sottolineano la **necessità di disporre di dati disaggregati** (per genere, età, disabilità, cittadinanza, condizione socioeconomica), aggiornati e tracciabili, accompagnati da metadati che ne esplicitino limiti, finalità e contesto di utilizzo. È stata inoltre evidenziata l'importanza di **distinguere tra dati strutturati** – tipicamente presenti nelle banche dati pubbliche – e **dati non strutturati**, spesso meno controllabili e più esposti a distorsioni. È emersa con chiarezza l'idea che la **“conformità tecnica” non sia sufficiente: accanto agli standard formali occorre integrare una valutazione etico-sociale della qualità del dato**. L'equità non è un attributo neutro, ma un concetto situato, che va valutato in relazione ai contesti, agli obiettivi perseguiti e agli impatti attesi.

Se i *bias* sono identificati come fenomeno strutturale, dalle sessioni emerge con forza che i **dati non sono neutri, i *bias* riflettono strutture culturali e narrative dominanti, e di conseguenza l'IA tende a replicare disuguaglianze preesistenti**.

Il linguaggio svolge un ruolo cruciale nell'addestramento dei sistemi di IA, perché i dati e i testi non neutri possono introdurre *bias* fin dall'origine. Un esempio è l'uso sistematico del maschile sovra-esteso nei testi e nelle basi informative, che può contribuire a strutturare dataset che assumono implicitamente il maschile come categoria di riferimento. Questo elemento linguistico, se non considerato criticamente, può influenzare i modelli di IA nella generazione di risposte, rafforzando associazioni e rappresentazioni sbilanciate. Altro esempio sono le *“master narratives”*, schemi culturali collettivi sedimentati. Se i dataset riflettono prevalentemente queste narrazioni dominanti, i modelli tenderanno a riprodurle come pattern “naturali”. Esempio classico: uomini associati a ruoli tecnici o scientifici, donne associate a ruoli di cura o relazionali. Per garantire risposte affidabili e inclusive occorre correggere le narrazioni dominanti, costruire narrazioni alternative (*counter-narratives*) e usare banche dati più rappresentative. Un punto chiave è aumentare la presenza di donne nei team di sviluppo, anche tramite quote femminili, per integrare prospettive diverse e ridurre il rischio del maschio come norma (*default male*).

Se da una parte non è realistico immaginare dataset completamente “fair” a monte, dall'altra diventa essenziale intervenire a valle, attraverso processi strutturati di auditing, pulizia, riequilibrio e documentazione dei dati, così da costruire dataset quanto più possibile equi e rappresentativi. In linea con gli orientamenti europei sull'IA affidabile e con l'approccio risk-based introdotto dall'AI Act, ciò implica l'adozione di valutazioni di impatto, meccanismi di monitoraggio continuo e strumenti di trasparenza (ad es. schede descrittive dei dataset e tracciabilità delle fonti), capaci di rendere espliciti i limiti, le lacune e le potenziali distorsioni. La *fairness*, dunque, non può essere presunta in origine, ma deve essere progettata, verificata e aggiornata nel tempo come parte integrante della governance dei sistemi di IA. Tra le proposte condivise si evidenzia dunque la necessità di un **monitoraggio e una valutazione continua degli output**. L'elemento interessante riguarda essenzialmente la proposta di un **coinvolgimento umano come “early detector” di *bias***.

Uno dei temi più strategici sui quali si è concentrata la discussione è il **ruolo della PA e del procurement**, dunque più in generale sul ruolo della PA nella governance pubblica dell'IA. La

PA è vista allo stesso tempo come: soggetto regolatore, committenza vigilante e custode dell'etica pubblica dell'innovazione.

In questo senso, le proposte principali sulle **azioni da suggerire** riguardano l'inserimento di **requisiti specifici nei capitolati** quando si acquisisce un software e si adotta un sistema; la **richiesta di valutazioni di conformità sui diritti fondamentali; audit algoritmici periodici; advisory board indipendenti; rafforzare la capacità negoziale verso fornitori privati; utilizzare la sandbox nazionale come ambiente controllato di sperimentazione.** Un significativo *insight* è che già la fase della raccolta dei dati che alimentano i sistemi riflette scelte politiche, non solo tecniche.

Un forte consenso si è registrato anche sulla necessità di avere **team interdisciplinari permanenti**, nello specifico rafforzare il coinvolgimento di esperti di genere, sociologi, giuristi, comunicatori; l'inclusione delle comunità rappresentate dai dati, ruolo dei Comitati Unici di Garanzia (CUG) nel monitoraggio, coinvolgimento della società civile. Questo punto rispecchia in sostanza anche l'idea chiave che muove il gruppo B5 nel 6NAP: l'equità non è un requisito tecnico isolato, ma richiede una mobilitazione e un coinvolgimento maggiore della componente civica per la difesa dei diritti.

Una delle criticità riscontrate riguarda il deficit culturale nella PA e una bassa consapevolezza su bias e processi algoritmici; si evidenzia la mancanza di una cultura diffusa del dato e la conseguente necessità di una maggiore **formazione proposta come leva strategica**, con percorsi continui e possibilmente pratici e laboratoriali, capaci di raggiungere anche i piccoli enti.

Tra i temi vi è quello della **promozione di modelli open source e, più in generale, della trasparenza e ispezionabilità dei sistemi.** È stata avanzata la proposta di **incentivare l'adozione di soluzioni aperte, o comunque verificabili, in grado di garantire replicabilità, verificabilità e maggiore controllo pubblico sui dataset e sui modelli utilizzati.** L'apertura e la spiegabilità sono considerate fattori potenzialmente abilitanti per rafforzare la fiducia dei cittadini e degli stakeholder, in quanto consentono una maggiore tracciabilità delle scelte tecniche e metodologiche. Allo stesso tempo, è emersa una consapevolezza condivisa: l'open source non equivale automaticamente ad equità. La trasparenza del codice o dei modelli, se non accompagnata da competenze adeguate e da una governance pubblica solida e competente, non garantisce di per sé la prevenzione dei bias o delle discriminazioni. L'apertura deve quindi essere integrata in un quadro più ampio di responsabilità istituzionale e capacità amministrativa.

Parallelamente, numerosi interventi hanno richiamato esplicitamente il quadro dei diritti fondamentali e il riferimento all'AI ACT. I partecipanti hanno sottolineato l'importanza della **valutazione di conformità e della valutazione d'impatto**, nonché dei *redress mechanisms* previsti dal regolamento, come strumenti essenziali per garantire la possibilità di tutela effettiva in caso di discriminazioni o usi impropri dei sistemi di IA. È stato evidenziato che la conformità non può essere intesa in senso meramente tecnico-formale, ma deve tradursi in

una **verifica sostanziale del rispetto dei diritti**, anche in chiave intersezionale. È inoltre emersa la necessità di un recepimento nazionale chiaro e operativo delle disposizioni europee, affinché i principi previsti dall'AI ACT diventino effettivamente praticabili. La convergenza tra i gruppi è stata netta su un punto: **l'equità dei dati non è un requisito accessorio, ma parte integrante della tutela dei diritti fondamentali e, come tale, deve essere garantita sin dalla fase iniziale di raccolta e validazione dei dati.**

Nel corso delle discussioni sono emerse dunque **criticità sistemiche** che ostacolano, in modo strutturale, la costruzione di dataset equi nella fase antecedente allo sviluppo degli algoritmi. Tra le principali barriere: la **carenza di dati disaggregati**, in particolare a livello locale, che limita la possibilità di analisi puntuali sulle disparità e riduce la capacità delle amministrazioni di progettare politiche realmente mirate. A ciò si aggiungono **fenomeni di frammentazione e scarsa interoperabilità tra banche dati**, che impediscono una visione integrata dei fenomeni e rendono complessa la costruzione di basi informative coerenti e comparabili.

Un ulteriore elemento critico riguarda la **persistente dominanza maschile nel settore tecnico e nello sviluppo dei sistemi digitali**, che può influenzare implicitamente scelte metodologiche e progettuali. Sono state inoltre richiamate **barriere culturali interne alla Pubblica Amministrazione**, legate alla difficoltà di adottare un approccio realmente interdisciplinare e orientato ai diritti.

In questo quadro, pesa anche **l'assenza di obblighi sanzionabili connessi all'adozione di linee guida e raccomandazioni**, che rischia di ridurre l'effettiva applicazione.

Infine, sono stati segnalati due fattori esterni rilevanti: **l'accesso limitato a dati detenuti da soggetti privati, spesso strategici per la comprensione dei fenomeni**, e il **digital divide**, che genera forme di invisibilità statistica per le fasce di popolazione meno digitalizzate, con conseguente rischio di esclusione sistemica.

Nonostante tali criticità, il livello di convergenza tra i gruppi è stato elevato su alcuni punti qualificanti. Vi è un consenso ampio sulla centralità della qualità dei dati come condizione preliminare dell'equità algoritmica, sulla necessità di una disaggregazione sistematica, sulla rilevanza della multidisciplinarietà e sulla centralità della formazione diffusa come leva di trasformazione culturale. È stato inoltre condiviso l'assunto che il monitoraggio debba essere continuo e strutturato e che la Pubblica Amministrazione debba esercitare un ruolo attivo, non meramente passivo, nella governance dei sistemi di IA. Fondamentale, infine, l'integrazione tra conformità tecnica e tutela sostanziale dei diritti fondamentali.

Tuttavia, alcune questioni richiedono un ulteriore approfondimento. Resta aperto il tema di **come rendere strutturale e possibilmente obbligatorio il monitoraggio senza appesantire eccessivamente i processi amministrativi**; come armonizzare il principio di minimizzazione dei dati previsto dal GDPR con l'esigenza di rappresentatività e disaggregazione; come strutturare advisory board realmente indipendenti e dotati di competenze adeguate; come definire standard nazionali condivisi per la disaggregazione dei dati; e, infine, come trasformare la formazione su open government, dati e IA in un requisito organizzativo stabile e non episodico.

L'analisi della prima sessione suggerisce che le future Raccomandazioni dovranno superare una visione meramente tecnica della *fairness*, esplicitando la dimensione politica, culturale e organizzativa della fase di raccolta e validazione dei dati. Dovranno inoltre rafforzare la funzione di committenza critica della Pubblica Amministrazione, prevedere meccanismi strutturati di monitoraggio e integrare in modo coerente formazione, governance e partecipazione come condizioni strutturali per garantire equità e tutela dei diritti fondamentali nello sviluppo e nell'adozione dei sistemi di intelligenza artificiale.

2.2 Sintesi del II incontro di co-creazione

Il secondo incontro di co-creazione ha affrontato il tema dell'**apertura del processo di progettazione, sviluppo e adozione degli algoritmi**.

Dai quattro gruppi di lavoro emerge che la consapevolezza è un elemento fondamentale per l'utilizzo dei sistemi di intelligenza artificiale (IA). I gruppi concordano che **il livello di consapevolezza attuale è disomogeneo e basso**, con conseguenze sulla capacità di valutare rischi e opportunità. Questa carenza riguarda sia i cittadini sia i dipendenti pubblici, che non hanno ancora sviluppato sufficiente capacità critica e tecnica per comprendere i sistemi di IA.

La formazione e l'alfabetizzazione digitale sono indicate come una risposta alla carenza di consapevolezza. L'alfabetizzazione dovrebbe essere trasversale tra pubblica amministrazione (PA) e cittadini, strutturale e non episodica. L'AI Act all'articolo 4 fornisce indicazioni sulla formazione, prevedendo obblighi per i fornitori di servizi. Auspicabile sarebbe **incrementare il monte ore obbligatorio annuale di formazione per la PA** e di introdurre moduli nei curricula scolastici e universitari. La formazione dovrebbe inoltre utilizzare campagne di informazione, iniziative rivolte a gruppi specifici e i punti di facilitazione digitale.

La trasparenza è individuata come condizione per costruire fiducia e garantire equità. Occorre distinguere tra trasparenza formale, che si riduce a un adempimento burocratico, e trasparenza sostanziale, che deve essere comprensibile e accessibile. Il diritto alla spiegazione delle scelte algoritmiche (*explainability* indicato nell'AI Act) e la verifica umana delle decisioni sono considerati elementi da garantire. Sul piano pratico, viene proposta l'adozione di algoritmi aperti verificabili, la richiesta del codice sorgente per le forniture pubbliche seguendo esempi come quelli dell'UNESCO, e la creazione di un simbolo che indichi l'uso dell'IA. È anche importante conoscere, dove possibile, i dati utilizzati nell'addestramento degli algoritmi, per individuare eventuali *bias* introdotti durante l'addestramento.

Il tema della governance ha una forte articolazione trasversale. I gruppi convergono sulla necessità di articolare responsabilità e strumenti su più livelli.

A livello nazionale, i gruppi indicano la necessità di linee guida cogenti per evitare che le amministrazioni si muovano in ordine sparso. Serve un orientamento comune con regole chiare che definiscano limiti costruttivi e promuovano consapevolezza. Viene sottolineata l'importanza di una autorità indipendente per la tutela del regolamento AI Act.

Sul piano operativo, la PA deve mantenere un ruolo attivo e non delegare completamente al privato lo sviluppo e la gestione delle soluzioni tecnologiche. La PA deve chiarire ruoli e responsabilità, favorire il reskilling dei dipendenti e valorizzare le competenze interne. Viene proposta l'introduzione di figure di garanzia come l'Algorithm Officer per il controllo e la verifica dei processi.

Il procurement pubblico, come già indicato nel report del gruppo 1, è indicato come strumento per orientare il mercato verso soluzioni eque e trasparenti. I gruppi concordano sulla necessità di clausole precise nei bandi di gara che vincolino i fornitori al rispetto di criteri di equità, trasparenza, accessibilità e riuso. L'Agenzia per l'Italia Digitale (AGID) ha pubblicato linee guida per la sicurezza nel procurement della PA; un approccio analogo andrebbe esteso all'IA. È in fase di stesura l'allegato 3 della Linea Guida 1 sull'adozione dell'IA per guidare la valutazione d'impatto. Il riuso di soluzioni validate attraverso piattaforme come developers.italia.it è visto come opportunità per ridurre costi e lock-in tecnologico (una situazione in cui un'organizzazione rimane vincolata a una determinata tecnologia, anche quando esistono alternative potenzialmente migliori), purché accompagnato da documentazione e governance.

La gestione dei dati è un fattore determinante. I dati dovrebbero essere equi e provenire da fonti attendibili: se provengono da enti omogenei del territorio che non dialogano tra loro, si crea il rischio di aumentare la disuguaglianza. I dataset devono essere statisticamente significativi e disaggregati per genere, età, provenienza e condizione socio-economica. Emerge la necessità di preferire l'elaborazione locale sui dispositivi quando possibile, di documentare quali dati vengono salvati e per quali finalità, e di applicare principi di minimizzazione, anonimizzazione e cancellazione by design.

Il monitoraggio costante è indicato come elemento chiave. Non è sufficiente valutare un algoritmo solo nella fase iniziale di adozione, servono controlli periodici condotti da gruppi multidisciplinari che includano informatici, tecnici, giuristi, sociologi, economisti, esperti di etica e rappresentanti della società civile. Viene proposta l'istituzione di un registro nazionale presso l'Autorità Nazionale Anticorruzione (ANAC) che raccolga informazioni su ogni appalto di soluzioni IA con dati minimi pubblicati in formato aperto per consentire controllo civico.

Il coinvolgimento della società civile è indicato come elemento trasversale. I comitati etici multidisciplinari e i tavoli permanenti di monitoraggio dovrebbero includere non solo tecnici

ma anche rappresentanti della società civile, esperti di diverse discipline e utenti vulnerabili. Sono citati esempi come il Board etico del Comune di Torino.

Sul piano della comunicazione, emerge la necessità di creare vocabolari comuni e figure ponte capaci di ridurre il divario tra sviluppatori e utenti finali. Il linguaggio specialistico rappresenta un ostacolo alla comprensione reciproca e alla traduzione dei bisogni in requisiti tecnici. In mancanza di questa mediazione, si rischia che i prodotti rispondano a criteri tecnici ma non ai valori delle comunità che li dovrebbero utilizzare. I gruppi hanno individuato come aree che richiedono particolare attenzione la sanità, dove l'IA in ambito diagnostico solleva interrogativi su responsabilità civili e penali, il welfare e i servizi essenziali, dove decisioni discriminatorie possono produrre effetti sui diritti delle persone, e la sorveglianza biometrica.

Emerge infine che **l'IA implica una scelta politica** e non è una questione solo tecnica ma anche sociale, che riflette scelte culturali e comporta responsabilità nella costruzione di un ecosistema tecnologico equo e inclusivo.

2.3 Sintesi del III incontro di co-creazione

Il terzo incontro di co-creazione “Sistemi di controllo e verifica dei sistemi di IA”, si è incentrato sulla fase di addestramento del sistema/strumento, per ridurre discriminazioni e stereotipi. L'incontro si è svolto il 7 ottobre 2025.

Durante la discussione gli aspetti più evidenziati sono stati i seguenti:

Trasparenza e dati di addestramento - La qualità e la provenienza dei dati utilizzati per l'addestramento sono prerequisiti fondamentali per garantire equità. È necessario rendere trasparenti: l'origine dei dataset, le modalità di raccolta, i settori della popolazione coinvolti, i criteri di selezione. **Si è suggerito di sviluppare un sistema di classificazione della maturità dei modelli, una sorta di “scoring” con livelli (A, B, C, D).** Ciò permetterebbe di distinguere, ad esempio, un sistema addestrato su milioni di dati e validato con criteri rigorosi da un modello in fase iniziale, con dataset limitati. Uno strumento del genere avrebbe un valore comunicativo importante, perché tradurrebbe parametri tecnici complessi in un linguaggio accessibile e utile per chiunque. L'AI Act trova una possibile risposta al problema della selezione dei dati, ponendo attenzione alla **formazione del team di addestramento**: il team, infatti, dovrebbe avere un'adeguata consapevolezza dell'importanza dei dati che vengono scelti e delle conseguenti necessarie azioni umane. Il team di designer dovrebbe includere dei componenti con competenze in materia di etica pubblica per intervenire sui temi che riguardano le discriminazioni e far sì che l'algoritmo le riconosca e le eviti. Per fare fronte a questo problema si possono usare i **presidi etici (commissioni) per fare dei test aperti in crowdsourcing (con persone con competenze specifiche sulle tematiche o meno).**

Serve documentare in modo chiaro le fasi di raccolta e trattamento dei dati, i criteri di selezione, gli algoritmi utilizzati e le modalità di verifica. È stata sottolineata la necessità di

rendere tracciabili le scelte metodologiche e di **mantenere un registro dei passaggi fondamentali dell'addestramento**, accessibile ai soggetti coinvolti nella governance e al controllo civico. La trasparenza deve riguardare anche le gare d'appalto e i capitolati pubblici, che dovrebbero integrare requisiti tecnici e procedurali mirati alla riduzione dei *bias* e alla verifica periodica dei risultati.

Addestramento responsabile, correzione delle distorsioni, *bias* - L'addestramento di un algoritmo è un processo iterativo, che richiede continui aggiustamenti. Dataset distorti, parziali o sbilanciati generano modelli con *bias* e scarsa capacità di generalizzazione. Molti squilibri presenti nei dati possono essere percepiti come "normali" e non problematici. Solo l'evoluzione culturale e sociale ha portato a riconoscerli come distorsioni. Questo significa che il riconoscimento dei *bias* non è neutro, ma dipende dal contesto storico e culturale. I dataset devono essere variegati, bilanciati e accuratamente puliti, rappresentando in modo equo i diversi gruppi sociali e culturali, validati con procedure *golden test* (test con una versione di riferimento considerata "d'oro", cioè perfettamente corretta e validata).

Necessario, pertanto, il coinvolgimento degli esperti di dominio (e lo scambio esperienziale con tecnici, tramite costituzione di gruppi di lavoro interdisciplinari) e etichettare le produzioni dell'IA in modo da renderle riconoscibili. Il sistema, inoltre, potrebbe essere addestrato a dubitare di sé stesso attraverso la *counter-narrative*: se si inseriscono nel sistema due tipi di indicatori diversi (*master-narrative* e *counter-narrative*), questi potrebbero nel tempo imparare a riequilibrarsi a vicenda. Nel dubbio, il sistema può essere allenato "a dire meno".

"Human in the Loop" - L'IA può evidenziare dati, ma la responsabilità finale deve rimanere al professionista umano, che interpreta e decide. Tuttavia permangono criticità: decidere di trascurare gli output, fare testing nel contesto reale, i modelli che si auto-modificano nel tempo, adattandosi continuamente ai nuovi input (come i GPAI).

Governance multilivello, segnalazioni, feedback - Tra le proposte arrivate, la possibilità di prevedere un soggetto centrale e un responsabile in ogni PA, supportato da un comitato etico interdisciplinare, che riunisca professionalità diverse e garantisca una valutazione più equilibrata e completa. **Necessario un meccanismo di feedback per segnalare la mancanza di trasparenza.**

Partecipazione attiva, consapevolezza diffusa, formazione - Fondamentale il ruolo degli attori della società civile, soprattutto qualora sia necessario fornire ai sistemi di IA documenti e pratiche di contesto che riflettano valori e istanze sociali. Tra le proposte operative: **istituire punti di contatto pubblici**, attraverso i quali la cittadinanza può segnalare malfunzionamenti o presentare reclami rispetto ai risultati prodotti dall'IA; abilitare piattaforme di consultazione aperta sui criteri di addestramento.

Fondamentale la **formazione capillare e interdisciplinare differenziata (verticalizzata) a seconda dei ruoli**: sviluppatori, dipendenti pubblici, decisori politici, aziende e cittadinanza devono acquisire competenze specifiche coerenti con le proprie responsabilità e anche con i propri obiettivi. La formazione alla quale si fa riferimento non è solo tecnica, ma anche culturale, finalizzata alla comprensione dei principi di equità, trasparenza e obbligo di render conto. Si tratta di un obiettivo di **alfabetizzazione collettiva alla cultura dell'IA**, che deve includere anche i giovani e i non addetti ai lavori. Consigliato l'uso di "mediatori linguistici" e di vocabolari condivisi.

Centrale anche la corretta comunicazione (anche per evitare il rischio manipolazione dell'informazione).

Differenze tra IA generativa e non generativa - Importante distinguere tra IA generativa e IA non generativa, evidenziando che i processi di addestramento e i rischi connessi ai bias sono diversi. L'IA generativa non punta a fornire risposte vere, ma "probabili" e questo la rende più vulnerabile a fenomeni di allucinazione o distorsione. Si osserva che, nella PA, la maggior parte dei sistemi attualmente in uso appartiene alla categoria non generativa (cioè sistemi che non creano nuovi contenuti ma prendono decisioni o effettuano previsioni sulla base di regole, modelli statistici o dati strutturati), che non deve essere dimenticata. Anche qui, il livello di spiegabilità e attendibilità dei modelli deve essere calibrato in base al contesto applicativo.

Normativa e decreti attuativi - In merito alla delega contenuta nella Legge n. 132 del 23 settembre 2025, e agli eventuali contenuti da includere nei decreti attuativi, è stata proposta la **creazione di un glossario normativo condiviso**. Le raccomandazioni che saranno prodotte nel percorso partecipativo possono contribuire a orientare la scrittura dei decreti attuativi, fornendo spunti concreti che tengano conto non solo delle esigenze istituzionali, ma anche delle prospettive della società civile e dei professionisti dei settori più coinvolti. La legge deve favorire un percorso di crescita culturale e linguistica che consenta a cittadini e organizzazioni di comprendere e gestire le tecnologie in modo critico. In questo senso, la normativa dovrebbe essere accompagnata da strumenti di educazione, sensibilizzazione e comunicazione trasparente.

Meccanismi di ricorso senza giudice - Previsti dall'AI Act europeo, consentirebbero, previa individuazione dell'autorità competente, la tutela dei diritti di cittadinanza digitale e procedure accessibili soprattutto per le categorie fragili.

Rischi legati alla localizzazione dei data center e alla sovranità dei dati - È emersa una preoccupazione rispetto alla localizzazione dei data center e alla dipendenza tecnologica da soggetti extraeuropei. Si teme la perdita di controllo sui dati e sull'uso che ne viene fatto al di fuori del contesto normativo europeo. **Si propone pertanto di privilegiare fornitori e partner**

europei, promuovendo un modello di autonomia strategica e di sovranità digitale, in linea con i principi del futuro AI Act e delle normative europee sulla protezione dei dati.

3. Prossimi passi verso la stesura delle Raccomandazioni

In linea con le principali raccomandazioni internazionali, l'interesse dei partecipanti ai tre incontri si è concentrato su aspetti specifici e ricorrenti aspetti.

Nella definizione della bozza di raccomandazioni, gli elementi e gli spunti emersi saranno clusterizzati sulla base dei [sette requisiti dell' *Trustworthy AI* elaborati dall'High-Level Expert Group on AI \(AI HLEG\)](#): intervento e sorveglianza umani; robustezza tecnica e sicurezza; privacy e governance dei dati; trasparenza; diversità, non discriminazione ed equità; benessere sociale e ambientale; responsabilità. Tali requisiti costituiscono un riferimento etico centrale nel quadro europeo sull'IA e risultano coerenti con l'impostazione dell'AI Act, che li traduce, in particolare per i sistemi ad alto rischio, in obblighi relativi alla qualità e governance dei dati, alla documentazione e tracciabilità, alla trasparenza, alla supervisione umana e alla robustezza, accuratezza e cybersicurezza.

Una volta completata la realizzazione e coordinamento del processo partecipativo con il supporto di People Powered e il coinvolgimento della community italiana di OGP IT e dei network attivi nell'Hub partecipazione e reso pubblico il Report sugli esiti del processo, il team di attuazione dell'Impegno B5 procederà alla "Stesura partecipata del documento di raccomandazioni al Governo sull'inclusività dei dati e degli algoritmi dell'IA", ovvero alla redazione di un testo in bozza che sarà sottoposto a consultazione pubblica online e che in seguito sarà trasformato in testo definitivo e promosso. Questa attività si concluderà con la pubblicazione di Raccomandazioni entro il mese di maggio 2026 e la trasmissione delle Raccomandazioni agli organismi nazionali competenti.

ITALIA

 OPEN GOV

Contributors:

Antonella Negri, Marieva Favoino, Leda Guidi, Giovanni Paolo Sellitto, Maria Morena Ragone, Fabio Riva, Francesca De Chiara, Iolanda Romano, Giulia Sudano, Elisabetta Todeschini, Glenda Gentili.

ITALIA

 OPENGOV



UNIONE EUROPEA
Fondo Sociale Europeo
Fondo Europeo di Sviluppo Regionale



Presidenza del Consiglio dei Ministri
Dipartimento per le politiche di coesione
e per il sud



Presidenza del Consiglio dei Ministri
Dipartimento della Funzione Pubblica

POC PROGRAMMA
OPERATIVO
COMPLEMENTARE

PNR GOVERNANCE
E CAPACITÀ
ISTITUZIONALE
2014-2020