POLLICINO & PARTNERS
AIDVISORY

La disciplina della trasparenza nell'AI Act: le nuove Linee guida della Commissione Europea sull'articolo 50

Oreste Pollicino[1], Rebecca Pupella[2], Sofia Menici[3]

1. Introduzione

Le Linee guida della Commissione Europea sull'articolo 50 dell'AI Act si collocano in una prospettiva che supera una concezione meramente formale della trasparenza. L'obiettivo non è aggiungere un obbligo informativo a valle dell'utilizzo dell'intelligenza artificiale, ma ricondurre la trasparenza a un principio strutturale di *governance*, destinato a permeare la progettazione, il funzionamento e la diffusione dei sistemi di AI. In questa prospettiva, essa diviene lo strumento attraverso cui rendere riconoscibili i rapporti di responsabilità che accompagnano la produzione e la circolazione di contenuti e decisioni generate o assistite dall'intelligenza artificiale.

Le soluzioni adottate dalle Linee guida riflettono tale impostazione, articolando gli obblighi di trasparenza in funzione dei diversi soggetti coinvolti e delle differenti modalità di impiego dell'intelligenza artificiale. Le Linee guida si inseriscono nel più ampio quadro di strumenti attuativi dell'articolo 50, comprendente anche il Code of Practice on Transparency of AI-Generated Content, riconosciuto dalla Commissione Europea quale strumento volontario che consente a *provider* e *deployer* di dimostrare la conformità agli obblighi di trasparenza.

Ne emerge un modello regolatorio nel quale la trasparenza costituisce uno strumento volto a preservare le condizioni di un'interazione consapevole con sistemi di intelligenza artificiale e, più in generale, la fiducia nell'ecosistema informativo digitale.

2. Analisi del contenuto

Le Linee guida sono articolate in nove sezioni:

(i) una prima sezione introduttiva che illustra contesto, finalità e la natura non vincolante del documento;

(ii) una seconda sezione che offre una panoramica degli obblighi di trasparenza previsti dall'articolo 50 e affronta alcune questioni orizzontali, tra cui: l'individuazione dei *provider* e dei *deployer* responsabili, l'ambito di applicazione dell'AI Act (sia con riferimento all'applicazione territoriale che con riferimento alla distinzione tra attività professionali e non), il rapporto con le pratiche vietate, i sistemi ad alto rischio e gli obblighi di *AI literacy* e quello con i modelli e i sistemi di IA per finalità generali;

---

[1] Managing Partner, Pollicino AIdvisory.
[2] Advisor AI Governance & Cybersicurezza, Pollicino AIdvisory.
[3] Dottoressa in Diritto e Tecnologia, Pollicino AIdvisory.

(iii) quattro sezioni che canalizzano separatamente i quattro obblighi di cui all'articolo 50: informazione delle persone che interagiscono direttamente con un sistema AI, marcatura e rilevabilità dei contenuti generati o manipolati dall'AI, informazione relativa ai sistemi di riconoscimento delle emozioni e di categorizzazione biometrica ed etichettatura dei *deep fake* e di determinati testi pubblicati su questioni di interesse pubblico;

(iv) una settima sezione che precisa i requisiti comuni alle informative, ossia chiarezza, distinguibilità, tempistica della comunicazione e accessibilità;

(v) un'ottava sezione che riguarda l'*enforcement*, il ruolo dei codici di buone pratiche, le autorità competenti, le sanzioni e l'entrata in applicazione degli obblighi; e

(vi) un'ultima sezione relativa a revisione e aggiornamento delle Linee guida.

Di seguito sono riepilogati i principali chiarimenti forniti dalla Commissione.

2.1. Sistemi AI che interagiscono direttamente con persone fisiche (art. 50, par. 1, AI Act)

Ai sensi dell'art. 50(1) AI Act, i *provider* di sistemi AI destinati a interagire direttamente con persone fisiche devono progettare e sviluppare tali sistemi in modo che le persone interessate siano informate di stare interagendo con un sistema AI.

La nozione comprende, tra gli altri, *chatbot*, assistenti vocali, agenti conversazionali, robot, avatar e agenti AI che interagiscono con persone nell'esecuzione di attività quali prenotazioni, gestione della corrispondenza, negoziazione, conclusione di contratti o acquisti.

Le Linee guida chiariscono che l'interazione deve avere carattere diretto e bidirezionale e che non rientrano normalmente nell'obbligo i sistemi che si limitano a raccogliere dati, selezionare o raccomandare contenuti, effettuare traduzioni o trascrizioni, assistere un operatore umano o comunicare esclusivamente con altre macchine.

Viene chiarito, poi, che l'informativa deve essere:

(i) fornita al più tardi al momento della prima interazione;

(ii) chiara e distinguibile;

(iii) inserita nel contesto concreto dell'interazione;

(iv) accessibile alle persone con disabilità;

(v) adattata quando il sistema può interagire con bambini, anziani o altri soggetti vulnerabili.

A tal fine, possono essere utilizzati banner, messaggi iniziali, dichiarazioni vocali, simboli visivi o combinazioni multimodali.

Non sono, invece, sufficienti, se utilizzate da sole:

(i) informazioni contenute nelle condizioni generali;

(ii) rinvii a URL o alla documentazione tecnica;

(iii) marcature percepibili soltanto dalle macchine;

(iv) formule generiche;

(v) indicazioni generali quali “questo sito utilizza l'AI”;

(vi) descrizioni meramente tecniche, quali “il sistema utilizza un LLM”.

Una comunicazione evidente prima della prima interazione è normalmente sufficiente, ma in contesti più sensibili o prolungati (a titolo esemplificativo, assistenza sanitaria, finanziaria o legale, gestione dei reclami, *AI companion* o interazioni con persone vulnerabili) possono essere necessari *reminder* periodici.

Chiarimenti specifici vengono, poi, forniti con riferimento agli agenti AI: le Linee guida precisano che il sistema dovrebbe comunicare sia la propria natura artificiale, sia la persona per conto della quale sta agendo. Tale obbligo non si applica quando la natura artificiale dell'interazione è evidente per una persona ragionevolmente informata, attenta e avveduta. La Commissione interpreta tuttavia questa eccezione in modo restrittivo, tenendo conto del pubblico previsto, delle persone ragionevolmente esposte, del livello di alfabetizzazione digitale e AI, del realismo dell'interfaccia, dell'avatar o della voce e dell'eventuale presenza di soggetti vulnerabili.

2.2. Marcatura e rilevabilità dei contenuti generati o manipolati dall'AI

Ai sensi dell'art. 50(2) AI Act, i *provider* di sistemi che generano o manipolano contenuti testuali, audio, immagini o video devono assicurare che gli output:

(i) siano marcati in un formato leggibile da macchina;

(ii) siano rilevabili come artificialmente generati o manipolati;

(iii) mediante soluzioni tecniche efficaci, interoperabili, robuste e affidabili.

Il documento chiarisce che la marcatura e la rilevabilità costituiscono due elementi distinti: garantire soltanto uno dei due non è sufficiente. Le tecniche utilizzabili possono comprendere *watermark*, metadati, *fingerprint*, sistemi crittografici, *logging* e strumenti per attestare la provenienza e l'autenticità dei contenuti.

Il *provider* può utilizzare una soluzione sviluppata da un fornitore a monte o da un soggetto terzo, ma conserva la responsabilità di dimostrare la conformità del proprio sistema.

Le Linee guida precisano che non tutti gli output prodotti attraverso un sistema AI devono essere marcati. Sono, tra gli altri, esclusi:

(i) la semplice riproduzione, selezione o organizzazione di contenuti già esistenti;

(ii) i dati registrati o trasmessi senza alterazioni;

(iii) alcune brevi sequenze di caratteri, etichette dell'interfaccia e contenuti analoghi;

(iv) il codice sorgente e determinate specifiche leggibili da macchina;

(v) le comunicazioni esclusivamente *machine-to-machine*;

(vi) le azioni e i passaggi intermedi degli agenti AI che non sono destinati a essere percepiti dalle persone;

(vii) gli interventi di editing standard e le modifiche che non alterano sostanzialmente l'input o il suo significato.

Le eccezioni devono essere valutate concretamente: correzioni grammaticali, conversioni di formato, compressione, riduzione del rumore o limitate modifiche di luce e colore possono rientrare nell'editing standard, mentre riassunti, riscritture sostanziali, sostituzioni di volti o voci e modifiche significative di persone, oggetti o eventi richiedono, in linea generale, la marcatura.

2.3. Sistemi di riconoscimento delle emozioni e categorizzazione biometrica

Ai sensi dell'art. 50(3) AI Act, i *deployer* di sistemi di riconoscimento delle emozioni o di categorizzazione biometrica devono informare le persone fisiche esposte che tali sistemi sono in uso.

Le Linee guida chiariscono che:

(i) l'informazione deve essere fornita al più tardi alla prima esposizione e può essere comunicata mediante avvisi testuali, comunicazioni orali, icone o combinazioni di tali strumenti;

(ii) l'obbligo di trasparenza non sostituisce gli ulteriori obblighi eventualmente applicabili in materia di protezione dei dati personali e non rende lecito un utilizzo vietato da altre disposizioni dell'AI Act;

(iii) è prevista un'eccezione per determinati impieghi consentiti dalla legge in materia di prevenzione, individuazione o investigazione di reati, nel rispetto delle garanzie applicabili.

2.4. Deep fake

I *deployer* che utilizzano professionalmente sistemi AI per generare o manipolare *deep fake* devono dichiarare, in modo percepibile dal pubblico, che il contenuto è stato artificialmente generato o manipolato.

La marcatura tecnica applicata dal *provider* ai sensi dell'articolo 50, paragrafo 2, non è sufficiente per adempiere all'obbligo del *deployer*: la *disclosure* deve essere direttamente visibile, udibile o altrimenti percepibile dalle persone.

I *deep fake* sono immagini, contenuti audio o un video generati o manipolati dall'AI che:

(i) presentano una somiglianza apprezzabile con persone, oggetti, luoghi, entità o eventi esistenti;

(ii) possono apparire falsamente autentici o veritieri.

La valutazione deve essere effettuata caso per caso, considerando il contesto di pubblicazione, il pubblico previsto e quello ragionevolmente prevedibile. Non è necessario dimostrare l'intenzione del *deployer* di ingannare.

Per *deep fake* che fanno evidentemente parte di opere artistiche, creative, satiriche, fittizie o analoghe si applica un regime attenuato: l'origine artificiale deve comunque essere dichiarata, ma con modalità che non ostacolino la normale fruizione dell'opera.

2.5. Testi pubblicati su questioni di interesse pubblico

I *deployer* devono dichiarare l'origine artificiale dei testi generati o manipolati dall'AI quando questi sono pubblicati allo scopo di informare il pubblico su questioni di interesse pubblico.

Secondo le Linee guida, possono rientrare in tale nozione, tra gli altri, i temi relativi a:

(i) politica e processi democratici;

(ii) amministrazione e servizi pubblici;

(iii) giustizia e diritti fondamentali;

(iv) salute e sicurezza pubblica;

(v) ambiente e sicurezza dei consumatori; e

(vi) sviluppi economici, finanziari, scientifici e culturali di rilevanza pubblica.

Non rientrano normalmente nell'obbligo la corrispondenza privata, le comunicazioni professionali individuali, i documenti interni e i contenuti accessibili soltanto a piccoli gruppi chiusi.

L'obbligo di *disclosure* non si applica quando ricorrono congiuntamente:

(i) un effettivo controllo umano o editoriale sul contenuto; e

(ii) l'assunzione della responsabilità editoriale da parte di una persona fisica o giuridica.

Viene chiarito che la revisione deve essere sostanziale e non meramente formale: non sono sufficienti correzioni ortografiche o grammaticali, controlli superficiali, approvazioni automatiche o revisioni effettuate esclusivamente da un altro sistema AI.

2.6. Individuazione dei soggetti responsabili

È opportuno distinguere tra:

(i) *provider*, responsabili principalmente della progettazione dei sistemi interattivi e della marcatura e rilevabilità tecnica dei contenuti;

(ii) *deployer*, responsabili principalmente dell'informazione relativa ai sistemi biometrici e dell'etichettatura percepibile di *deep fake* e testi di interesse pubblico.

Un'organizzazione è generalmente *deployer* quando assume la responsabilità della decisione di utilizzare il sistema, delle sue finalità e delle modalità concrete d'impiego, anche se non ne controlla tecnicamente il funzionamento.

Viene chiarito che:

(i) i dipendenti e i collaboratori che operano sotto le istruzioni e il controllo dell'organizzazione non sono normalmente considerati *deployer* autonomi;

(ii) gli intermediari che si limitano a ospitare, trasmettere o distribuire contenuti prodotti da terzi, senza avere autorità sul sistema utilizzato per generarli, non sono *deployer* ai sensi dell'AI Act. Sono tuttavia incoraggiati a preservare le marcature e le etichette lungo la catena di distribuzione.

Resta inteso che lo stesso soggetto possa assumere contemporaneamente il ruolo di *provider* e *deployer*, con conseguente applicazione cumulativa dei relativi obblighi.

2.7. Esclusioni e ambito territoriale

L'AI Act, ai sensi dell'art. 2 dello stesso, non si applica alle persone fisiche che utilizzano un sistema AI esclusivamente per finalità personali e non professionali. Tale esclusione riguarda, però, soltanto gli obblighi che gravano su chi utilizza il sistema, ossia il *deployer*: non esonera, invece, il *provider* dagli obblighi che gli competono.

Le Linee guida indicano, tra gli esempi di uso personale, anche la generazione di *deep fake* destinati a essere condivisi sui social media. Ciò non significa però che tali contenuti siano automaticamente leciti: restano applicabili le altre norme europee e nazionali eventualmente rilevanti.

Sono inoltre previste esclusioni specifiche per alcune attività di ricerca e sviluppo. Quanto ai sistemi open source, il fatto che un sistema sia distribuito con licenza libera non lo sottrae agli obblighi dell'articolo 50, se tali obblighi risultano applicabili. Diversamente, i singoli componenti open source che, da soli, non costituiscono un sistema AI non sono direttamente soggetti a tali obblighi.

Infine, gli obblighi dell'articolo 50 possono riguardare anche soggetti stabiliti fuori dall'Unione europea, quando il sistema o i suoi output sono utilizzati nell'UE. Tuttavia, secondo le Linee guida, un utilizzo nell'Unione del tutto occasionale, imprevedibile o non autorizzato non è, da solo, sufficiente per assoggettare un *provider* extra-UE che non abbia immesso il sistema sul mercato europeo né lo abbia messo in servizio nell'Unione.

2.8. Requisiti comuni delle informative

Viene chiarito che tutte le informazioni previste dall'articolo 50 devono essere fornite:

(i) in modo chiaro e distinguibile;

(ii) al più tardi alla prima interazione o esposizione; e

(iii) nel rispetto dei requisiti di accessibilità applicabili.

Il riferimento alla prima interazione o esposizione riguarda ciascuna persona interessata. Per i contenuti, la *disclosure* deve quindi accompagnare l'output nei confronti del pubblico che vi è esposto. Quando è prevedibile che una persona possa iniziare a visualizzare un contenuto dopo il suo avvio, una *disclosure* mostrata soltanto all'inizio potrebbe non essere sufficiente e dovrebbe essere integrata con indicazioni successive o persistenti.

2.9. Codici di buone pratiche ed enforcement

Provider e *deployer* soggetti agli obblighi relativi ai contenuti generati o manipolati dall'AI possono dimostrare la conformità aderendo a un codice di buone pratiche valutato come adeguato ai sensi dell'articolo 50, paragrafo 7 dell'AI Act.

Resta possibile utilizzare mezzi alternativi, ma l'operatore dovrà essere in grado di documentare e dimostrare l'adeguatezza delle misure adottate.

3. Considerazioni conclusive

Nel complesso, la disciplina delineata dalle Linee guida conferma l'approccio sistemico sotteso all'articolo 50 dell'AI Act. Particolarmente significativa, in tale prospettiva, è la disciplina degli agenti AI, chiamati a dichiarare non soltanto la propria natura artificiale, ma anche per conto di chi agiscono: alla trasparenza dell'identità si affianca così quella della delega e, dunque, della responsabilità. La medesima impostazione si riflette nella disciplina dei contenuti sintetici, nella quale gli obblighi di marcatura tecnica gravano sui *provider*, mentre quelli di *disclosure* percepibile ricadono sui *deployer*, con un'eccezione per i testi di interesse pubblico sottoposti a un controllo umano realmente sostanziale e riconducibili a una precisa responsabilità editoriale.