IL CODICE DI CONDOTTA PER L'INTELLIGENZA ARTIFICIALE GENERALE

17 luglio 2025

AUTORI

Aurora Agostini
Partner

Giulietta Minucci
Counsel

Giovanni Lombardi
Associate

Alessandro Carlini
Associate

L'intelligenza artificiale generativa ha ormai raggiunto una maturità tale da richiedere un intervento normativo specifico e articolato. In questo contesto, l'Unione Europea ha sviluppato un approccio pionieristico attraverso l'adozione del Codice di Condotta per i modelli di intelligenza *artificiale general-purpose* ("GPAI")[1], strumento giuridico che si inserisce nel più ampio quadro Regolamento UE 2024/1689 ("AI Act"), e che rappresenta un *unicum* nel panorama regolamentare internazionale.

Il presente contributo si propone di analizzare dal punto di vista giuridico le implicazioni, le opportunità e le criticità di questo innovativo *framework* normativo, esaminando in particolare i tre pilastri fondamentali del Codice – trasparenza, sicurezza e diritto d'autore – con un *focus* trasversale sulla gestione della responsabilità sistemica per i modelli che presentano rischi elevati. Tale analisi risulta particolarmente rilevante considerando che il Codice, pur mantenendo natura volontaria, costituisce il principale strumento attraverso cui i fornitori di modelli GPAI possono dimostrare la propria conformità agli obblighi previsti dall'AI Act.

CONTESTO NORMATIVO

I modelli GPAI sono soggetti a obblighi specifici delineati negli **articoli 53 e 55** dell'AI Act: l'articolo 53 stabilisce obblighi per tutti i fornitori di modelli GPAI, mentre l'articolo 55 introduce ulteriori requisiti per i modelli che presentano "rischi sistemici" (definiti nell'articolo 3, paragrafo 65, come rischi specifici delle capacità ad alto impatto che hanno effetti significativi sul mercato dell'Unione).

[1] I modelli di intelligenza artificiale *general-purpose* (GPAI) sono sistemi addestrati su *dataset* ampi e diversificati, capaci di generare output versatili e di adattarsi a una molteplicità di compiti senza essere progettati per una funzione specifica. Tra i principali esempi di GPAI figurano GPT-4 di OpenAI, Gemini di Google, Claude di Anthropic e Stability Diffusion di Stability AI. Questi modelli costituiscono la base tecnologica per un'ampia gamma di applicazioni downstream, che spaziano dai chatbot aziendali agli assistenti virtuali, dagli strumenti di traduzione automatica ai sistemi di analisi dei dati, fino a soluzioni creative per la generazione di immagini e contenuti audiovisivi. La loro diffusione capillare, unita alla capacità di incidere su settori strategici e sulla sfera dei diritti fondamentali, conferisce loro un potenziale impatto sistemico sull'economia e sulla società nel suo complesso.

Il Codice non è un mero atto di autoregolamentazione settoriale, bensì lo strumento di ***soft-law*** espressamente previsto e incentivato dall'articolo 56: tale disposizione disciplina la "*elaborazione, adozione e approvazione*" di codici di condotta volti a garantire la corretta applicazione degli obblighi che gravano sui fornitori di GPAI (artt. 53 e 55).

L'adesione al Codice è formalmente volontaria: tuttavia, l'articolo 56(1) chiarisce che esso costituisce il "*principale strumento*" attraverso cui i *provider* possono dimostrare la propria conformità; in assenza di adesione, il fornitore è onerato di provare – "con mezzi alternativi adeguati" – il rispetto puntuale di tutti gli obblighi, con evidente aggravio probatorio e reputazionale.

L'articolo 56(2) identifica quattro aree che un codice deve obbligatoriamente coprire:

- mantenimento aggiornato della documentazione tecnica (art. 53(1)(a)-(b));

- definizione del grado di dettaglio adeguato del training-data summary (art. 53(1)(d));

- mappatura delle fonti di rischio sistemico dei GPAI nell'Unione;

- procedure di assessment e mitigazione proporzionate di tali rischi.

Il Code of Practice assolve puntualmente a queste quattro funzioni attraverso i suoi capitoli "Trasparenza", "Copyright" e "Safety & Security".

IL CAPITOLO TRASPARENZA: DOCUMENTAZIONE E *ACCOUNTABILITY*

GLI OBBLIGHI DOCUMENTALI

Il Capitolo Trasparenza del Codice rappresenta la concretizzazione operativa degli obblighi previsti dall'articolo 53, paragrafi 1(a) e 1(b), dell'AI Act: si tratta di un insieme di impegni – denominati *Measures* – cui i fornitori di GPAI aderenti al Codice si obbligano per assicurare che la documentazione tecnica dei modelli sia non solo completa e aggiornata, ma anche accessibile agli stakeholder rilevanti.

Queste ***Measures*** non sono meri adempimenti formali: esse costituiscono un vero e proprio ***framework*** di ***accountability***, finalizzato a garantire che tutti gli attori della catena del valore – dalle autorità di vigilanza fino ai *downstream providers* (i "fonitori a valle" della versione italiana del Regolamento) – dispongano delle informazioni necessarie per integrare, valutare e monitorare i GPAI in modo conforme agli *standard* europei di sicurezza e tutela dei diritti fondamentali.

Tale documentazione è destinata a tre categorie di soggetti:

- AI Office (AIO): destinatario delle informazioni più dettagliate, fornite su richiesta;
- Autorità nazionali competenti (NCAs): destinatarie di informazioni specifiche per l'esercizio delle funzioni di vigilanza;
- Fornitori *downstream* (DPs): destinatari delle informazioni necessarie per l'integrazione dei modelli nei propri sistemi AI.

Questa tripartizione riflette il principio di proporzionalità: le informazioni rese proattivamente ai downstream providers sono di natura generale e funzionale all'integrazione del modello, mentre quelle destinate alle autorità vengono fornite solo in risposta a richieste formalizzate, con l'obbligo di indicare la base giuridica e lo scopo del trattamento.

Sul punto, vale la pena di evidenziare che una delle sfide centrali del Capitolo Trasparenza è la tutela dei segreti commerciali e delle informazioni riservate: l'articolo 78 AI Act impone ai destinatari delle informazioni (AIO, NCAs, DPs) l'obbligo di rispettare la riservatezza dei dati ricevuti, adottando adeguate misure di cybersecurity per proteggerne la confidenzialità

Measure 1.1: redazione e aggiornamento della documentazione

La prima Measure obbliga i firmatari a predisporre, già al momento dell'immissione del modello sul mercato, una documentazione tecnica esaustiva che copra ogni aspetto significativo del GPAI. Tale documentazione, raccolta nel Model Documentation Form, include informazioni sull'identità del fornitore, le caratteristiche architetturali del modello, le specifiche tecniche e progettuali, i processi di addestramento (con dettagli sulle metodologie utilizzate e le logiche di design), i dati trattati (tipologie, provenienza, metodologie di cura, misure per il rilevamento di bias e contenuti inadeguati) e infine i consumi energetici e computazionali.

La natura dinamica di questa Measure richiede che la documentazione sia costantemente aggiornata per riflettere modifiche o aggiornamenti del modello, con l'obbligo di conservare anche le versioni precedenti per un periodo di dieci anni.

Measure 1.2: comunicazione delle informazioni agli stakeholder

La seconda *Measure* riguarda la messa a disposizione delle informazioni contenute nel Model Documentation Form verso tre distinti destinatari:

- AI *Office* e Autorità nazionali competenti, che possono richiedere l'accesso a informazioni dettagliate per esercitare le loro funzioni di supervisione, sempre nel rispetto del principio di necessità e proporzionalità;
- *downstream providers*, che devono poter accedere alle informazioni rilevanti per comprendere le capacità e i limiti del modello al fine di integrarlo responsabilmente nei propri sistemi di IA.

Questa comunicazione deve avvenire entro termini ragionevoli, e mai oltre 14 giorni per le richieste dei *downstream providers*, salvo circostanze eccezionali.

Measure 1.3: qualità, integrità e sicurezza della documentazione

La terza *Measure* impone ai firmatari di garantire che le informazioni documentate siano non solo accurate, ma anche protette contro alterazioni non intenzionali e accessi non autorizzati. A tal fine, essi sono incoraggiati ad adottare protocolli di qualità e standard tecnici consolidati, rafforzando così la fiducia nella robustezza e integrità dei dati condivisi..

BILANCIAMENTO TRA TRASPARENZA E SEGRETI COMMERCIALI

Il Capitolo Trasparenza affronta una delle tensioni più delicate per i fornitori di GPAI: la necessità di garantire un elevato livello di trasparenza verso autorità e *partner* commerciali senza compromettere la riservatezza delle informazioni strategiche. L'articolo 78 dell'AI Act impone ai destinatari di tali dati di rispettare rigorosi obblighi di confidenzialità e di implementare misure di sicurezza informatica adeguate per proteggere diritti di proprietà intellettuale e segreti commerciali.

Il Codice adotta un approccio modulare e calibrato: le informazioni da fornire proattivamente ai downstream providers sono limitate a ciò che è strettamente necessario per consentire una integrazione sicura e conforme del modello; le informazioni più sensibili destinate ad AI Office e NCAs vengono invece comunicate solo su richiesta motivata e circoscritta agli elementi indispensabili per l'esercizio delle funzioni di vigilanza. Questo sistema riflette l'intento dell'Unione Europea di creare un equilibrio dinamico tra trasparenza e protezione della competitività industriale, favorendo al contempo un ecosistema di IA più affidabile e responsabile..

IL CAPITOLO *COPYRIGHT*: POLITICHE, TUTELE E RESPONSABILITÀ

L'OBBLIGO DI UNA *COPYRIGHT POLICY*

Il Capitolo Copyright del Codice rappresenta la risposta operativa all'articolo 53, paragrafo 1(c), dell'AI Act, che impone ai fornitori di GPAI immessi nel mercato dell'Unione di adottare una politica per assicurare la conformità al diritto UE in materia di copyright e diritti connessi. Tale obbligo nasce dalla necessità di garantire che i modelli di IA non siano alimentati da contenuti protetti in violazione delle normative vigenti e che i risultati generati non producano a loro volta atti di violazione.

La politica sul *copyright* prevista dal Codice non si limita a principi astratti, ma definisce azioni concrete che i fornitori devono intraprendere per mappare, monitorare e mitigare i rischi legati all'utilizzo di contenuti tutelati, secondo un approccio proporzionato alla dimensione e alle risorse del provider.

LE MISURE DEL CAPITOLO *COPYRIGHT*: UN *FRAMEWORK* DI DILIGENZA

Measure 1.1 – Redigere, aggiornare e attuare una *copyright policy*

I firmatari si impegnano a elaborare e mantenere aggiornata una politica aziendale sul copyright che disciplini le modalità con cui i GPAI vengono addestrati e utilizzati nel rispetto dei diritti d'autore. Questa policy deve individuare le responsabilità interne per la sua attuazione e prevedere meccanismi di verifica e controllo, divenendo così un elemento chiave della governance interna in materia di proprietà intellettuale. Inoltre, i fornitori sono incoraggiati a pubblicare una sintesi della loro policy per aumentare la trasparenza verso *stakeholder* esterni.

Measure 1.2 – Accesso lecito ai contenuti protetti

Il Codice richiede ai fornitori di garantire che l'estrazione di dati e contenuti dal web tramite crawling avvenga solo rispetto a materiali cui si ha accesso lecito. Ciò comporta il divieto di aggirare misure tecnologiche efficaci (es. paywall o subscription models) e l'obbligo di escludere dai processi di crawling quei siti web riconosciuti dalle autorità come ripetutamente violativi del copyright su scala commerciale

Measure 1.3 – Identificazione e rispetto delle riserve di diritti

In linea con l'articolo 4(3) della Direttiva (UE) 2019/790, i firmatari devono adottare tecnologie all'avanguardia – comprese soluzioni machine-readable come il protocollo robots.txt – per individuare ed escludere contenuti per i quali i titolari dei diritti abbiano espresso riserve di utilizzo ai fini del *text and data mining*. Tale misura sottolinea il principio secondo cui il rispetto delle riserve di diritti è un elemento essenziale della legittimità del processo di *training*.

Measure 1.4 – Prevenzione delle violazioni nei risultati generati

Non meno rilevante è la misura che impone di prevenire, tramite salvaguardie tecniche, la generazione di output da parte dei GPAI che riproducano in modo illecito contenuti protetti. A ciò si aggiunge l'obbligo di includere nei termini di utilizzo del modello, o nella documentazione di accompagnamento per i modelli open source, il divieto esplicito di usi che violino i diritti d'autore.

Measure 1.5 – Punti di contatto e gestione dei reclami

Infine, i firmatari devono designare un punto di contatto elettronico per i titolari dei diritti e predisporre un meccanismo per la ricezione e gestione dei reclami

relativi a presunte violazioni. Tale misura mira a garantire un canale diretto di dialogo con i soggetti lesi e a rafforzare l'accountability dei fornitori verso l'ecosistema creativo

TEMPLATE PER IL RIEPILOGO DEI DATI DI ADDESTRAMENTO

Parallelamente al Codice, l'AI Office sta sviluppando un template standardizzato per il riepilogo dei contenuti di addestramento che i fornitori devono rendere pubblicamente disponibile ai sensi dell'articolo 53, paragrafo 1(d), dell'AI Act. Questo template si articola su tre sezioni principali:

1. **informazioni generali**: identificazione del modello e del fornitore, date di immissione sul mercato, caratteristiche generali dei dati di addestramento;
2. **elenco delle fonti di dati**: categorizzazione dettagliata delle fonti (dataset pubblici, dati di terze parti, dati crawled, dati sintetici) con specifiche indicazioni dimensionali e temporali;
3. **aspetti rilevanti del trattamento dei dati**: misure per il rispetto del copyright, rimozione di contenuti indesiderati, altri aspetti rilevanti del processing.

IL CAPITOLO SICUREZZA E GESTIONE DEI RISCHI SISTEMICI: UNA *GOVERNANCE* DEL RISCHIO *FUTURE-PROOF*

UN FRAMEWORK INTEGRATO PER I GPAI AD ALTO RISCHIO

Il Capitolo Sicurezza del Codice di Condotta dà attuazione concreta all'articolo 55 dell'AI Act, che impone ai fornitori di GPAI con rischio sistemico di istituire una governance tecnica e organizzativa in grado di monitorare, valutare e mitigare i rischi lungo tutto il ciclo di vita del modello.

Questa disciplina non si esaurisce in un insieme di precetti astratti, ma prende la forma di **dieci** ***commitments*** interconnessi, ognuno accompagnato da misure operative pensate per guidare i firmatari verso l'adozione di pratiche allo stato dell'arte..

UNA STRUTTURA MODULARE E PROPORZIONATA

I dieci commitments rappresentano i pilastri su cui poggia il Safety and Security Framework dei fornitori GPAI:

- Alcuni sono di natura strategica, come l'obbligo di stabilire procedure di governance e di designare responsabili interni per la gestione dei rischi.
- Altri hanno un taglio tecnico-operativo, imponendo misure di sicurezza informatica, sistemi di monitoraggio post-market e piani di risposta agli incidenti.
- Infine, alcuni commitments promuovono una cultura aziendale del rischio, attraverso formazione, audit interni e meccanismi di revisione continua.

L'obiettivo complessivo è duplice: da un lato garantire che i GPAI non diventino vettori di minacce sistemiche per la salute, la sicurezza pubblica o i diritti fondamentali; dall'altro rafforzare la resilienza dei fornitori di fronte a un panorama di minacce in continua evoluzione.

LE PRINCIPALI MISURE: UNA NARRAZIONE DELLA DILIGENZA TECNICA E ORGANIZZATIVA

Senza ridurre la ricchezza del *framework* a un elenco statico, è possibile evidenziare alcune aree chiave:

- Governance del rischio (Commitments 1-3): prevedono la creazione di politiche interne, la designazione di un responsabile per la sicurezza e la definizione di criteri per la valutazione dei rischi.
- Valutazione e mitigazione (Commitments 4-6): includono l'obbligo di condurre analisi dei rischi sistemici, utilizzare tecniche avanzate di testing (es. red-teaming) e adottare misure correttive per mantenere i rischi entro livelli accettabili.
- Sicurezza operativa (Commitments 7-9): disciplinano la protezione contro minacce esterne e interne, la sicurezza fisica e informatica delle infrastrutture, e la continuità operativa.
- Accountability e trasparenza (Commitment 10): impongono la redazione di un Safety and Security Model Report, da condividere con l'AI Office e da pubblicare in forma sintetica per informare il pubblico, bilanciando trasparenza e tutela dei segreti commerciali.

VALUTAZIONI ESTERNE INDIPENDENTI

Un elemento innovativo del Codice riguarda l'obbligo di fornire accesso a valutatori esterni indipendenti per facilitare il monitoraggio post-market. I firmatari devono fornire a un numero adeguato di valutatori esterni indipendenti accesso gratuito e adeguato a:

- Le versioni del modello più capaci riguardo al rischio sistemico;
- Le chain-of-thought del modello, se disponibili;

- Le versioni del modello con il minor numero di mitigazioni di sicurezza implementate.

Tale accesso può essere fornito attraverso API, accesso on-premise, accesso tramite hardware fornito dal firmatario, o rendendo pubblicamente disponibili i parametri del modello per il *download*.

PROFILI CRITICI, PROSPETTIVE EVOLUTIVE E CONSIDERAZIONI SISTEMATICHE

Il Codice di Condotta presenta una contraddizione che non può essere ignorata: si proclama volontario, ma nella pratica funziona come se fosse obbligatorio. L'articolo 56 dell'AI Act ha creato un meccanismo ingegnoso quanto insidioso: identifica il Codice come principale strumento probatorio per dimostrare la conformità agli obblighi normativi, trasformando quello che dovrebbe essere un atto di buona volontà in una necessità commerciale. Il risultato è paradossale: pur non essendo giuridicamente vincolante, l'adesione al Codice diventa praticamente inevitabile. Chi si sottrae deve affrontare l'onere probatorio alternativo di dimostrare la conformità attraverso altre modalità - un percorso inevitabilmente più costoso, complesso e dall'esito incerto. È una forma di coercizione soft, elegante ma efficace.

Questa configurazione solleva interrogativi scomodi sulla legittimazione democratica. Stiamo assistendo alla creazione di normative sostanzialmente cogenti elaborate al di fuori dei tradizionali circuiti decisionali democratici. Il processo multi-stakeholder, per quanto tecnicamente sofisticato e rappresentativo, non può supplire alla legittimazione che deriva dal procedimento legislativo ordinario. Si profila un vulnus democratico che richiede un controllo parlamentare più stringente sui contenuti del Codice e sui suoi futuri aggiornamenti.

Ma le criticità non si fermano qui. Il Codice si trova intrappolato in una missione quasi impossibile: conciliare l'inconciliabile. Da un lato, titolari di diritti d'autore e organizzazioni della società civile chiedono trasparenza totale; dall'altro, le imprese tecnologiche devono proteggere segreti commerciali e investimenti miliardari in ricerca e sviluppo. È una quadratura del cerchio che il legislatore europeo ha tentato di risolvere con un equilibrismo normativo dai risultati incerti. Il Model Documentation Form rappresenta l'emblema di questa tensione: la differenziazione tra informazioni destinate ai diversi stakeholder cerca di accontentare tutti, ma rischia di non soddisfare nessuno. Il risultato potrebbe essere doppiamente fallimentare: troppo opaco per garantire quella trasparenza che la società civile legittimamente rivendica, troppo invasivo per preservare quella competitività industriale che le imprese considerano vitale.

Ancora più problematico è il framework per la gestione dei rischi sistemici. Come si può definire ex ante soglie di accettabilità per tecnologie la cui evoluzione procede per salti discontinui e imprevedibili? I "systemic risk tiers" previsti dal Codice rischiano di trasformarsi in esercizi di compliance puramente formali, incapaci di intercettare quei rischi emergenti che, per definizione, sfuggono alle attuali metodologie di valutazione. È come tentare di prevedere il futuro con gli strumenti del passato.

L'assenza di sanzioni dirette per la violazione degli impegni assunti non significa assenza di conseguenze giuridiche - tutt'altro. Si configura un sistema di enforcement indiretto che opera su multiple dimensioni, creando una rete di pressioni sottili ma pervasive. L'enforcement reputazionale presuppone un mercato in cui i consumatori siano capaci di valutare e penalizzare comportamenti non conformi - ipotesi difficilmente verificabile nel mercato B2B dei modelli GPAI, dove gli acquirenti sono spesso altrettanto sofisticati tecnologicamente quanto i fornitori. L'enforcement contrattuale dipende dalla capacità negoziale dei soggetti coinvolti e potrebbe rivelarsi un'arma spuntata nei rapporti caratterizzati da significativo squilibrio di potere. Infine, l'enforcement regolamentare richiede un'attività di vigilanza continua e specializzata da parte dell'AI Office, il cui successo dipenderà dalle risorse umane e tecniche che verranno effettivamente allocate - una variabile tutt'altro che scontata.

Un aspetto sottovalutato ma potenzialmente esplosivo riguarda le implicazioni per il diritto della concorrenza europea. La standardizzazione di pratiche operative attraverso il Codice potrebbe facilitare comportamenti collusivi o comunque restrittivi della concorrenza, particolarmente pericolosi in un mercato già caratterizzato da elevata concentrazione. L'obbligo di fornire accesso a valutatori esterni indipendenti potrebbe inoltre creare asimmetrie informative tra incumbent e nuovi entranti, consolidando le posizioni dominanti esistenti. È un effetto collaterale che potrebbe trasformare uno strumento di regolazione in un meccanismo di protezione delle rendite di posizione.

Paradossalmente, le previsioni in materia di trasparenza e documentazione potrebbero favorire l'emergere di nuovi operatori specializzati in servizi di compliance e valutazione, creando nuove dinamiche competitive. L'equilibrio tra questi effetti contrapposti dipenderà largamente dalle modalità applicative del Codice e dalle scelte interpretative dell'AI Office - un potere discrezionale considerevole che merita attento monitoraggio.

Il Codice rappresenta un banco di prova per un nuovo modello di regolazione dell'innovazione tecnologica, che abbandona l'approccio tradizionale command-and-control in favore di un sistema basato su principi, obiettivi e processi. È un esperimento affascinante ma rischioso: pur garantendo maggiore flessibilità e adattabilità all'evoluzione tecnologica, solleva interrogativi fondamentali circa la certezza del diritto e la prevedibilità delle conseguenze giuridiche. La sfida principale consiste nel mantenere un adeguato livello di determinatezza normativa

pur consentendo l'adattamento continuo ai cambiamenti tecnologici - un equilibrio che richiede finezza di calibrazione quasi chirurgica.

L'approccio europeo è destinato a influenzare significativamente l'evoluzione della regolamentazione internazionale dell'intelligenza artificiale. La *leadership* europea, consolidata dall'*early mover advantage* dell'AI Act, potrebbe favorire l'emergere di standard globali basati sui principi e metodologie sviluppati in ambito UE. Tuttavia, la convergenza regolamentare internazionale dovrà confrontarsi con approcci normativi significativamente diversi: quello statunitense basato sulla self-regulation e quello cinese caratterizzato da forte controllo statale. Il successo del modello europeo dipenderà dalla sua capacità di dimostrarsi efficace nel bilanciare innovazione e protezione dei diritti, evitando al contempo effetti negativi sulla competitività delle imprese europee.

In definitiva, il Codice di Condotta rappresenta un esperimento regolamentare di portata storica, il cui successo dipenderà dalla capacità del sistema di dimostrare concretamente la propria efficacia nel perseguire gli obiettivi dichiarati, senza compromettere l'innovazione tecnologica né creare indebite barriere competitive. Sarà cruciale lo sviluppo di meccanismi di monitoraggio e valutazione che consentano di misurare oggettivamente l'impatto del Codice e di apportare le necessarie correzioni di rotta.

AUTORI

Aurora Agostini

Partner

Giulietta Minucci

Counsel

Giovanni Lombardi

Associate

Alessandro Carlini

Associate