La responsabilità nella value chain dell'AI

Lidia

Adaptive Legal Assistant

Introduzione

In occasione del lancio di **Lidia 2.0**, nuova piattaforma di intelligenza artificiale per il settore legale presentata a Milano il **18 giugno 2025,** l'avv. **Marco Pagani** ha riunito – nel ruolo di moderatore – tre esperte di primo piano per discutere la **Responsabilità nella value chain dell'AI.** Il panel ha offerto un confronto fra normativa, prassi industriale e best practice.

- **Sabrina Chibbaro** - Notaia presso AC Notai - presso ha aperto i lavori intrecciando il paradigma **risk-based** dell'AI Act con l'esigenza di una **vigilanza umana continua.** Secondo la relatrice, la mera compliance procedurale agli artt. 9-13 del Regolamento non basta a inoculare i bias strutturali dei dataset: serve un presidio operativo che consenta al professionista di contestare l'output deviato;

- La Dott.ssa **Anna Lanza** - Manager of compliance AI presso Fastweb - ha poi introdotto il modello dei **«cerchi concentrici» di responsabilità,** provider del modello, provider di piattaforma, deployer-studio legale, cliente-committente, rappresentando l'asimmetria regolatoria che oggi grava sul terzo anello. Per riequilibrare la filiera ha proposto un modello di responsabilità più diffusa che contempli tutta la value chain dell'AI.

- Infine, l'avv. **Emanuela Semino** - co-founder and partner presso Vectis - ha analizzato la responsabilità in relazione alla **diligenza professionale**, che parte dall'**AI literacy**, e ai presidi tecnici dei fornitori e all'explainability dei risultati. Solo rendendo comprensibili processi e le fonti utilizzate è possibile consentire allo studio legale di presidiare la propria sfera di responsabilità e di trasformare i rischi dell'IA in fattori governabile.

Pur muovendo da angolature diverse, le tre relatrici convergono su un punto cardine: la responsabilità lungo la catena dell'IA non è un evento residuale, ma un **processo di qualità continua** che richiede **governance basata sul rischio, sorveglianza umana qualificata e trasparenza tecnica.** I contributi che seguono forniscono la mappa concettuale per declinare tali principi nei contesti professionali a più alto impatto. Nelle **conclusioni** presentate al termine degli interventi, LIDIA propone un quadro sintetico delle sfide aperte e delle soluzioni operative emerse dal dibattito.

Responsabilità e intelligenza artificiale: tra risk-based approach e vigilanza operativa (Sabrina Chibbaro)

La responsabilità giuridica nei sistemi di IA non può ridursi a un mero elenco di obblighi ex-ante: serve intrecciare il paradigma risk-based dell'AI Act (Reg. UE 2024/1689) con una vigilanza umana continua, così da evitare che gli errori strutturali di dataset e modelli si riversino sull'operatore di ultima istanza, cioè il professionista. «La mera compliance procedurale non è uno scudo assolutorio» ha avvertito **Sabrina Chibbaro**, ricordando che chi si limita a soddisfare gli articoli 9-13 del Regolamento può comunque lasciare irrisolti i difetti che generano bias o allucinazioni .

1. Il cuore normativo dell'AI Act e i suoi limiti

L'AI Act adotta un modello di market access che:

- divide i sistemi in quattro **fasce di rischio** (inaccettabile, alto, limitato, minimo);
- impone, per quelli ad alto rischio, gestione ciclica del rischio, dataset di alta qualità, documentazione tecnica e **sorveglianza umana "effettiva e continua"** (art. 14).

Eppure – osserva Chibbaro – questi requisiti restano preventivi: non disciplinano il nesso di causalità né il regime della colpa. Se la sorveglianza umana si riduce a un kill switch o a un "sì/no" finale, «l'umano diventa il capro espiatorio dei difetti della macchina» . Serve quindi un controllo operativo che consenta di contestare l'output quando il modello devia dal mandato, evitando il "colpo di spugna" sulla responsabilità del fornitore.

2. Il ponte con il diritto italiano

Il **Disegno di legge S. 1146/2024** mira a raccordare l'AI Act con l'ordinamento interno, inserendo:

- **Principio di controllo umano**: ogni sistema di IA deve operare sotto supervisione professionale tracciabile;
- **Clausola per le libere professioni**: l'IA può solo supportare – mai sostituire – l'attività intellettuale; la violazione costituisce illecito deontologico, con sanzioni pecuniarie fino a 5 milioni di euro – una norma nata proprio per scongiurare lo scenario denunciato da Chibbaro di un professionista "sorvegliato" ma deresponsabilizzato ;
- **Sandbox nazionali** vigilate da AgID e Garante Privacy per sperimentazioni ad alto impatto;
- **Programmi obbligatori di AI literacy** per PA e professioni regolamentate, in linea con l'idea che «l'uso dell'IA non ci autorizza a rimanere ignoranti» .

Il DDL si affianca inoltre alla Direttiva UE 2024/2853 (prodotto difettoso digitale) e alla proposta di AI Liability Directive, che introduce una presunzione di causalità quando si violano gli obblighi dell'AI Act.

3. Verso un modello ibrido: rischio, vigilanza e trasparenza explainable-by-design

Le indicazioni emerse nel panel convergono su tre direttrici:

- **Governance risk-based:** rating interni, audit periodici, registri d'uso accessibili alle autorità.
- Vigilanza umana qualificata: l'operatore deve poter interrogare il modello, ricostruire la catena logica e rigettare l'output in caso di deviazioni; una validazione binaria degrada il controllo

controllo e trasferisce la colpa su chi convalida.

- **Trasparenza tecnica e RAG:** l'integrazione di meccanismi Retrieval-Augmented Generation (RAG) consente di visualizzare le fonti e ridurre le allucinazioni. Chibbaro ricorda che i fornitori devono «istruire l'utente» e adottare guardrail di qualità delle fonti; la trasparenza abilita la responsabilità lungo tutta la filiera .

Ogni fase – addestramento, fine-tuning, rilascio, impiego – richiede quindi una **responsabilità condivisa**: dal curatore dei dataset al professionista che usa la piattaforma, passando per il fornitore che implementa i guardrail.

4. Conclusione

Integrare risk governance e human vigilance significa trasformare la responsabilità da "evento residuale" a processo di qualità continua. Il quadro europeo offre la griglia; il disegno di legge italiano ne precisa il perimetro settoriale e pone paletti dove l'AI Act tace. Ma solo l'innesto di audit robusti, formazione permanente e piattaforme explainable-by-design – secondo la visione di Sabrina Chibbaro – potrà prevenire le distorsioni: bias, allucinazioni e confusione normativa. In assenza di tali presìdi, la promessa dell'IA di potenziare la professione rischia di trasformarsi in un moltiplicatore di contenziosi, con la colpa che, anziché risalire al difetto algoritmico, finirebbe addossata all'ultimo anello della catena. Una responsabilità trasparente, monitorata e condivisa resta dunque la chiave per coniugare innovazione e tutela dei diritti fondamentali nell'ecosistema dell'intelligenza artificiale.

La filiera dell'IA: verso una responsabilità "a cerchi concentrici" (Anna Lanza)

«La trasparenza abilita la responsabilità» — con questa sintesi **Anna Lanza** ha aperto il suo intervento, denunciando l'asimmetria dell'AI Act: su 113 articoli, soltanto uno disciplina in modo diretto i doveri del model provider, mentre la gran parte del carico regolatorio grava sullo studio legale che integra l'LLM nel proprio workflow. Da qui l'idea dei **quattro cerchi concentrici di responsabilità** — fornitore del modello, fornitore della piattaforma, deployer-professionista e cliente-committente — ognuno chiamato a presidiare un segmento distinto del rischio algoritmico.

1. Il modello di base: zone d'ombra

Per Lanza il primo anello è il LLM-provider. I suoi impegni — selezione di fonti attendibili, documentazione trasparente del training e safety report pubblico — derivano dagli articoli 9-13 dell'AI Act, che impongono registri di testing e convalida dei dataset. Tuttavia, una volta che il modello approda all'**open market** il produttore perde il controllo dei contesti d'uso, che potrebbero includere applicazioni ad "alto" o "inaccettabile" rischio ai sensi del Regolamento (EU) 2024/1689. Per colmare il vuoto, Lanza invoca la condivisione dei log di prova e dei dati di addestramento con chiunque intenda fine-tunare o integrare l'LLM: «il dovere di trasparenza corre lungo tutta la catena, o diventa un alibi di irresponsabilità».

2. Il fornitore di piattaforma: da fine-tuning a guardrail

Il secondo cerchio spetta al **platform-provider**, responsabile di adattare il modello al dominio giuridico e di inserirvi guardrail tecnici (filtri, rate limit, prompt shield) insieme a un testing «valido e ripetibile». Un flusso di prompt che funziona dieci volte, ricorda Lanza, può fallire all'undicesima: la probabilità resta la natura dell'LLM. Questo presidio operativo trova riscontro:

- nello **ISO/IEC 42001:2023**, primo standard sui management system per l'IA, che impone audit ciclici e miglioramento continuo;
- nel **NIST AI RMF – Generative AI Profile**, che raccomanda il post-market monitoring dei modelli generativi e la gestione dei rischi lungo tutto il ciclo di vita.

Così il platform-provider diventa «amministratore di contesto»: sceglie l'LLM più idoneo, lo addestra sul corpus normativo, documenta i failure mode e allerta il deployer quando la probabilità d'errore supera la soglia di guardia.

3. Il deployer-studio legale: oltre la figura del "terzo incolpevole"

Terzo cerchio: lo **studio legale** che integra l'IA non può rifugiarsi nel ruolo di end-user passivo. L'AI Act gli impone human oversight «effettiva e continua» (art. 14) e introduce il principio di **AI literacy** come prerequisito di tutti gli obblighi. Lanza porta l'esempio di **Fastweb**, che ha istituito una rete decentrata di "AI champion" interni, affiancata a un Codice di Condotta che vieta l'upload di dati sensibili nei tool pubblici e impone disclaimer sul rischio residuo. Compiti chiave del deployer:

- valutazione integrata dei rischi e data protection impact assessment;

- formazione del personale e up-skilling giuridico-tecnico;
- validazione manuale degli output (human-in-the-loop), con possibilità e dovere di rigettare le risposte spurie;
- informazione trasparente al cliente sull'impiego dell'IA.

---

4. Il committente: dalla passività alla vigilanza contrattuale

Nel quarto cerchio Lanza colloca il cliente-committente, tradizionalmente percepito come mero destinatario del servizio. Al contrario, egli deve partecipare alla due diligence dell'IA con:

- **clausole di audit** sugli algoritmi impiegati, accesso ai log e obblighi di notifica degli incidenti;
- **divieti d'uso improprio e penali** in caso di re-prompting fraudolento;
- indennizzi per violazioni di diritti fondamentali.

La rinnovata **Product Liability Directive 2024/2853** include software e sistemi di IA nella nozione di prodotto e introduce presunzioni di difetto e di nesso causale, alleggerendo l'onere probatorio dell'utente. A ciò si aggiunge la proposta di **AI Liability Directive**, che presume la causalità quando il fornitore o la piattaforma violano gli obblighi del Regolamento. Ne deriva un cliente più attivo, che pretendendo accountability contrattuale esercita un ruolo di controllo diffuso.

---

5. Verso una catena di responsabilità distribuita

Il modello "a cerchi concentrici" di Lanza non è mera teoria, ma un framework operativo che intreccia:

- **standard tecnico-organizzativi** (ISO 42001, NIST AI RMF);
- **obblighi regolatori** (AI Act, PLD, futura AILD);

- **formazione diffusa** (AI literacy);
- **clausole contrattuali di accountability.**

Se ognuno presidia il proprio cerchio — qualità e trasparenza per il provider, guardrail e testing per la piattaforma, risk governance e verifica professionale per il deployer, vigilanza contrattuale per il cliente — il rischio algoritmico da mina vagante si trasforma in fattore governabile. «La catena si spezza — conclude Lanza — solo quando un anello finge che il rischio appartenga all'anello successivo». Integrare questi livelli significa, in ultima analisi, rendere l'IA compatibile con il dovere di diligenza professionale e con la tutela effettiva dei diritti fondamentali.

Dalla AI literacy all'explainability: la catena della diligenza (Emanuela Semino)

«Stiamo consegnando ai professionisti uno strumento potentissimo, ma pericoloso se non usato bene» ricorda l'avvocata Emanuela Semino, invitando gli studi legali a presidiare l'intero ciclo di vita dell'IA, «dalla prima riga di prompt all'ultima firma sul documento» . La sua proposta di "catena della diligenza" capovolge l'idea dello studio come «terzo incolpevole» e distribuisce il rischio su quattro livelli — provider di modello, piattaforma verticale, professionista, cliente — ognuno dotato di compiti di competenza, controllo e tracciabilità.

1. AI literacy come prerequisito professionale

Per Semino la formazione è il primo argine al rischio: «Lo studio deve trasformare i propri avvocati in esperti dello strumento che decidono di usare» . Non basta un corso frontale: serve un up-skilling continuo sulle potenzialità e, soprattutto, sui limiti degli LLM, perché «nove volte su dieci l'errore nasce da un prompt sbagliato». Il richiamo alla AI literacy — rilanciato anche da altre relatrici come Anna Lanza — ricalca l'impostazione dell'AI Act, che impone a provider e deployer un presidio costante delle competenze interne.

2. Explainability & RAG by design: vedere (e validare) le fonti

«L'LLM non ragiona: genera testi probabilistici», avverte Semino. Per fidarsi occorre poter ricostruire il "percorso di pensiero". Da qui la richiesta di architetture Retrieval-Augmented Generation che affianchino all'output atti, sentenze e norme pertinenti, così da rendere immediata la verifica di coerenza . L'explainability non è un optional cosmetico: riduce il rischio di "contratti Frankenstein" e tutela il cliente in caso di contenzioso.

3. Testing validato e ripetibile: l'audit come guard-rail operativo

Le piattaforme legali, sostiene Semino, devono «mettere in piedi un processo di testing valido e replicabile» per i prompt-template che mettono a disposizione degli utenti . Un flusso che funziona dieci volte può fallire all'undicesima per il comportamento stocastico dell'LLM; perciò servono test regressivi a ogni release, log degli errori e piani di rollback.

4. Responsabilità diffusa: istruzioni chiare, tracce indelebili

Il professionista resta il presidio critico, «ma la piattaforma ha l'obbligo di spiegare come usare l'IA» . Se l'utente ignora le guide, carica documenti irrilevanti o usa template sbagliati, l'errore è dietro l'angolo. Per dimostrare l'human oversight richiesto dall'art. 14 AI Act, Semino consiglia di documentare prompt, revisioni e correzioni: la "paper trail" diventa prova di diligenza.

5. Potenziamento, non sostituzione

«L'IA è un potenziamento, non un rimpiazzo del giudizio umano» — concetto ribadito più volte nel panel e fatto proprio da Semino . La vigilanza deve abbracciare scelta del modello, ingestione dei dati, stesura dei prompt, revisione dell'output e aggiornamento normativo: se manca un aggiornamento 231, l'avvocato deve accorgersene prima di consegnare l'atto. La diligenza non si delega alla macchina.

Conclusione

Formare le persone, rendere spiegabili le macchine, tracciare ogni decisione: su questi tre pilastri Semino costruisce una governance che trasforma l'IA generativa in un potenziamento sicuro e responsabile. L'avvocato resta garante della qualità, la piattaforma fornisce i guard-rail e il cliente partecipa con una due-diligence informata. Una responsabilità distribuita, documentata e monitorata che converte il rischio algoritmico da mina vagante a fattore governabile.

Conclusioni

Il confronto fra le tre relatrici converge su un principio irrinunciabile: la r**esponsabilità lungo la value chain dell'intelligenza artificiale è un processo continuativo di qualità,** non un rimedio ex-post. Dal dibattito emergono tre coordinate operative che riannodano, senza strappi, l'introduzione del panel con i contributi di dettaglio.

Governance basata sul rischio, sorvegliata e dimostrabile
Il paradigma dell'AI Act – fasce di rischio, obblighi di gestione ciclica e sorveglianza umana "effettiva e continua" – traccia il perimetro normativo; il Disegno di legge S. 1146/2024 lo declina nel contesto professionale italiano, imponendo controllo umano tracciabile e sandbox vigilate. Ne discende l'esigenza di audit periodici, registri d'uso accessibili e metriche di affidabilità che traducano la compliance formale in presidio sostanziale.

Sorveglianza umana qualificata e AI literacy diffusa.
Chibbaro lo definisce "presidio operativo", Lanza lo inquadra nel terzo cerchio di responsabilità, Semino lo eleva a prerequisito deontologico: in tutti i casi, il professionista resta l'ultimo garante. Programmi obbligatori di formazione, reti di AI champion interni e documentazione puntuale di prompt e revisioni costituiscono la prova della vigilanza richiesta dall'art. 14 AI Act.

Trasparenza tecnico-giuridica come collante della filiera
Explainability e architetture RAG, invocate da tutte le relatrici, permettono di ricostruire il percorso logico dell'LLM, riducendo bias e allucinazioni e fornendo la paper trail essenziale nei contenziosi. Log condivisi, guardrail replicabili e clausole contrattuali di audit saldano tra loro i quattro cerchi di responsabilità – modello, piattaforma, studio, cliente – evitando che il rischio si sposti, incontrollato, da un anello all'altro.

Lidia

+39 010 8991141
info@lidiatech.ai
www.lidiatech.ai
/company/lidiatech