

MODULO 11 • AGOSTO 2026

Profili Giuridici ed Etici dell'AI Forense

Ammissibilità, spiegabilità, bias e responsabilità:
ciò che decide se il lavoro tecnico regge in aula

Al termine del modulo lo studente sarà in grado di

01 • ARGOMENTARE

L'ammissibilità e il peso di risultati AI alla luce dei criteri di scientificità della prova e della ripetibilità processuale

02 • INQUADRARE

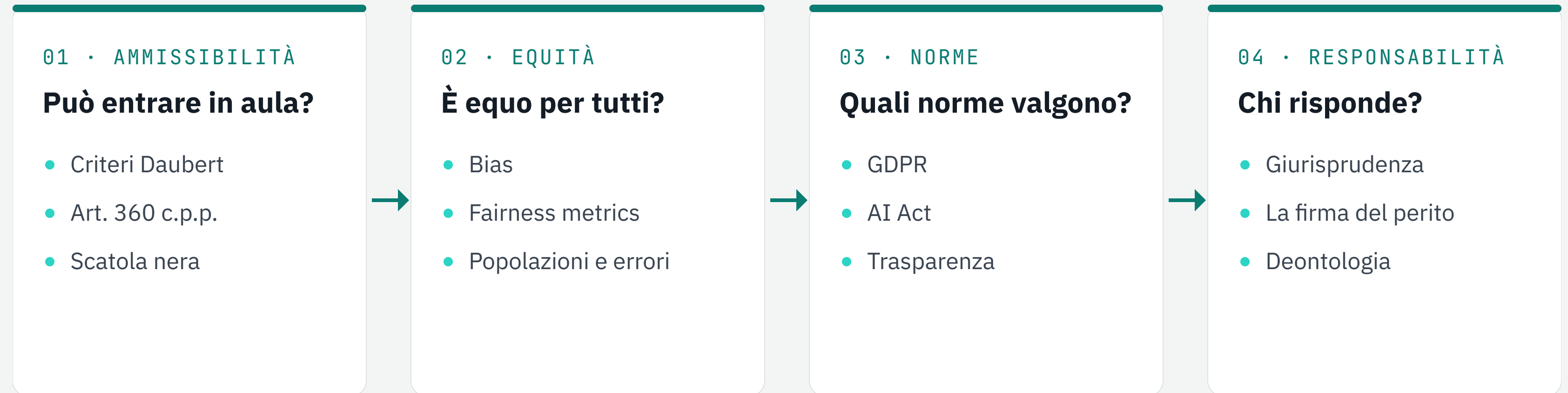
Il trattamento dei dati nell'indagine AI-assistita entro GDPR e AI Act, con le relative basi giuridiche e cautele

03 • ASSUMERE

La responsabilità professionale delle conclusioni fondate su strumenti algoritmici, resistendo all'automation bias

Il diritto chiede al perito di conoscere e dichiarare quanto e dove l'AI sbaglia, senza delegarle il giudizio. Le regole del modulo applicano questo principio.

Le quattro domande del giurista all'AI forense



Le quattro aree si susseguono nelle slide del modulo nello stesso ordine dello schema: ciascuna presuppone la precedente.

La scientificità del metodo va dimostrata

I **criteri Daubert** (ordinamento statunitense, riferimento comparatistico): verificabilità del metodo, peer review, **tasso di errore noto**, standard di esecuzione, accettazione nella comunità scientifica. Le metriche del modulo 7 ne sono la traduzione tecnica.

Nell'ordinamento italiano il nodo è la **ripetibilità** (artt. 359–360 c.p.p.) e il libero convincimento motivato del giudice: un output non deterministico e non riproducibile (modello aggiornato, temperatura non fissata) mette in tensione il contraddittorio tecnico.

Opacità e motivazione

Il perito deve mettere il giudice in condizione di motivare: la **explainable AI** (modulo 7) e la documentazione del modello (dataset, versione, validazione) ne sono il presupposto minimo.

Il solo rinvio all'esito dell'algoritmo non costituisce motivazione.

Per ogni strumento AI impiegato la relazione riporta **cinque voci obbligatorie**: identificazione e versione, funzione svolta, metriche dichiarate e loro fonte, protocollo di validazione applicato, ruolo dell'esito nel ragionamento peritale.

Cinque domande a un detector di deepfake

Criterio	Applicato a un detector di deepfake impiegato in perizia
Verificabilità e falsificabilità	Esistono benchmark pubblici su cui chiunque possa misurarne le prestazioni?
Peer review e pubblicazione	Architettura e valutazioni sono pubblicate in sedi scientifiche, o il prodotto è opaco?
Tasso di errore noto	FPR e FNR sono misurati su dati comparabili al reperto (compressione, durata, generatore)?
Standard di esecuzione	Esiste un protocollo d'uso (preprocessing, soglie)? Il perito lo ha seguito e documentato?
Accettazione nella comunità	La tecnica è riconosciuta dalla letteratura forense, o è una novità isolata?

ESEMPIO • Detector A: pubblicato, benchmark pubblici, FPR 4% su video compressi; il reperto però è un audio di 8 secondi, fuori dal dominio validato. Detector B: proprietario, «accuratezza 99%» da brochure, nessun dettaglio. Il perito usa A dichiarandone il limite di dominio e affiancandolo all'esame tradizionale; esclude B, carente su quattro criteri su cinque.

L'art. 360 c.p.p. e l'output non deterministico

Molti sistemi di AI producono output **non deterministici** (temperatura, aggiornamenti del modello, servizi remoti): se la difesa ripete l'analisi con esito diverso, l'accertamento perde attendibilità.

Per questo occorre **ingegnerizzare la ripetibilità**:

- Modello in locale o versione bloccata; mai endpoint che evolvono senza controllo
- Parametri fissati (temperatura 0 ove possibile, seed dichiarato)
- Log integrale di input, configurazione e output, con hash
- Conservazione dell'ambiente (container o immagine della workstation d'analisi)

ESEMPIO • DUE QUALIFICAZIONI

Caso ripetibile: classificazione di immagini con modello locale a versione fissata, seed e log conservati; un terzo ottiene gli stessi esiti: art. 359.

Caso irripetibile: nel 2024 un'analisi fu condotta su un servizio esterno oggi dismesso; l'esatto esito non è più riproducibile. Il valore residuo sta tutto nella documentazione contestuale; per questo quella scelta si sconsiglia a monte.

La ripetibilità va **costruita prima dell'analisi**: dichiararla a posteriori, nella memoria difensiva, non basta.

Errori sistematici e regole che li presidiano

Bias algoritmici

Errori **sistematici per gruppi**: riconoscimento facciale meno accurato per alcune popolazioni, NLP che degrada su dialetti e gerghi, classificatori che riflettono squilibri del dataset.

In sede forense il bias colpisce persone concrete: indagati, vittime, terzi.

GDPR

Base giuridica del trattamento, **minimizzazione** e proporzionalità (richiamo al modulo 2), categorie particolari di dati, limiti alle decisioni unicamente automatizzate, sicurezza del trattamento in laboratorio.

AI Act

I sistemi AI per **law enforcement** rientrano tra gli alti rischi: obblighi di gestione del rischio, qualità dei dati, documentazione tecnica, sorveglianza umana, accuratezza e robustezza. Pratiche vietate per i casi estremi.

Le tre cornici **convergono**: chiedono di conoscere e documentare i limiti dello strumento e di lasciare la decisione a una persona.

Fairness metrics: misurare il bias

L'imparzialità di un sistema va misurata. Le metriche principali confrontano i gruppi su cui il sistema opera: **demographic parity** (stessi tassi di esito positivo), **equalized odds** (stessi tassi di falso positivo e falso negativo), **calibrazione per gruppo** (lo stesso score significa la stessa probabilità in ogni gruppo).

Quando i gruppi hanno prevalenze diverse, queste proprietà sono **matematicamente incompatibili**: un sistema non può soddisfarle tutte. Scegliere quale garantire è quindi una **decisione di valore**, da prendere in modo esplicito e motivato.

Chiedere o produrre le metriche di errore disaggregate per gruppo fa parte della due diligence del perito su qualunque strumento AI usato nell'indagine.

ESEMPIO • IL RICONOSCIMENTO DISEGUALE

Un sistema di comparazione facciale dichiara un tasso di falso positivo complessivo dell'1%. Disaggregato per gruppi demografici, il tasso varia di un ordine di grandezza, come documentato dagli studi NIST FRVT su molti sistemi commerciali.

Per un indagato del gruppo più penalizzato il «1%» dichiarato non descrive il rischio di errore reale. In controesame va quindi chiesto *su quale popolazione* sia stato misurato il tasso d'errore.

Il trattamento dei dati nell'indagine AI-assistita

Analizzare reperti significa trattare dati personali, spesso di terzi estranei ai fatti. I principi che governano il perito:

- **Base giuridica:** il penale segue il regime della direttiva law enforcement (in Italia, d.lgs. 51/2018); in ambito civile e aziendale, il GDPR con le sue basi (obbligo legale, legittimo interesse bilanciato)
- **Minimizzazione e proporzionalità:** si tratta ciò che il quesito richiede: è il richiamo all'acquisizione selettiva motivata del modulo 2
- **Categorie particolari** (salute, orientamenti, biometria): trattamento da motivare specificamente
- **Sicurezza:** reperti cifrati, accessi tracciati, laboratorio segregato; **mai dati su servizi AI esterni** non controllati, che configurerebbero una comunicazione a terzi priva di base

ESEMPIO • IL TERZO NELLA CHAT

Nell'estrazione del dispositivo dell'indagato compaiono conversazioni intime con una persona estranea ai fatti e dati sanitari di un familiare.

Condotta corretta: il materiale estraneo al quesito non entra in relazione né negli allegati; se tecnicamente separabile, resta fuori dalla pipeline NLP; la sua esistenza è annotata solo ove necessario alla completezza. Il criterio è la pertinenza al quesito.

L'AI Act per il perito: rischio, obblighi, divieti

Il regolamento europeo gradua gli obblighi sul **rischio**. Gli impieghi in ambito law enforcement e giustizia rientrano tipicamente tra gli **alti rischi**, con obblighi che per il perito si traducono in requisiti familiari:

- Sistema di gestione del rischio e **qualità dei dati** di addestramento
- **Documentazione tecnica** e registrazione degli eventi (log)
- **Sorveglianza umana** effettiva (lo human-in-the-loop del modulo 8)
- **Accuratezza e robustezza** dichiarate (le metriche del modulo 7)

Per i casi estremi il regolamento prevede divieti, come l'identificazione biometrica remota in tempo reale negli spazi pubblici, salvo eccezioni tassative.

Avvertenza: dopo l'omnibus sull'AI (Reg. (UE) 2026/1744) gli obblighi per l'alto rischio dell'Allegato III si applicano dal 2 dicembre 2027; lo stato puntuale va verificato alla data dell'incarico.

ESEMPIO • LA DOMANDA IN CAPITOLATO

Un laboratorio valuta uno strumento AI di triage: oltre alle prestazioni, chiede al fornitore la classificazione di rischio ai sensi dell'AI Act, la documentazione tecnica e le modalità di sorveglianza umana previste.

Se il fornitore non sa rispondere, il laboratorio si espone a problemi di conformità e a obiezioni in controesame.

La prova algoritmica nei tribunali

Il problema

Sulla prova generata o selezionata dall'AI la giurisprudenza è **giovane e frammentaria**, con pochi precedenti consolidati. Il quadro citato può cambiare fra il deposito e l'udienza.

I punti relativamente fermi

L'output algoritmico non si auto-autentica: serve un **testimone o perito che ne risponda**. La «scatola nera» indebolisce il contraddittorio. Custodia, ripetibilità e motivazione valgono come per ogni strumento.

La prassi prudente

Datare ogni affermazione giuridica («allo stato attuale»); distinguere norma vigente, orientamento e dottrina; **fondare la consulenza sul metodo tecnico** documentato.

ESEMPIO • Un consulente fonda l'ammissibilità del proprio strumento su una pronuncia di merito isolata; in appello la pronuncia è superata e la relazione perde credito. Un collega fonda la stessa conclusione sul metodo (validazione, tasso d'errore, riproducibilità) e cita la giurisprudenza solo a conferma: la sua relazione resta valida.

La firma in calce resta una: quella del perito

Il perimetro della delega

Lo strumento può cercare, ordinare e suggerire; valutare la rilevanza, attribuire condotte e trarre conclusioni spetta al perito. Chi trascrive un output non verificato ne risponde come di un errore proprio.

Automation bias

La fiducia eccessiva nell'esito automatico è documentata e cresce con la stanchezza e il carico. Contromisure strutturali: validazione in cieco (modulo 8), revisione tra pari, tempo dedicato ai casi discordanti.

Doveri di trasparenza

Dichiarare l'uso di AI anche se non richiesto; conservare log e versioni per il controesame; segnalare i limiti noti dello strumento, compresi quelli che indeboliscono la propria tesi.

Le domande tipiche della difesa in controesame: con quali dati è stato addestrato lo strumento? Qual è il tasso di errore su dati come quelli del caso? Chi ha verificato l'esito, e come? La relazione deve già rispondere a ciascuna.

Codici deontologici e certificazioni

I codici deontologici di settore condividono alcuni nuclei: **competenza dichiarata** (nessun incarico fuori dal proprio perimetro), **imparzialità** (il consulente serve l'accertamento, anche quando è di parte), **riservatezza, integrità del metodo** (nessuna conclusione oltre i dati), **trasparenza sugli strumenti**, compresi l'uso di sistemi AI e i loro limiti.

Le **certificazioni** (ISO/IEC 17025 per i laboratori, percorsi individuali per i professionisti), pur non richieste dalla legge italiana, pesano sulla credibilità del consulente. L'uso di AI aggiunge il dovere di **mantenere la competenza** necessaria a capire e spiegare gli strumenti impiegati.

Chi firma deve poter difendere in controesame ogni riga della relazione: l'uso dell'AI non trasferisce la responsabilità (slide precedente).

ESEMPIO • LA RELAZIONE «SCRITTA DALL'AI»

Un consulente usa un LLM per redigere parti della relazione. In controesame emergono una citazione normativa inesistente e un riferimento bibliografico inventato.

L'uso dello strumento è legittimo se dichiarato e verificato. Il vizio sta nella firma senza verifica, perché ogni riga depositata è del firmatario; ne esce compromessa la credibilità dell'intera consulenza, comprese le parti solide.

Controesame simulato su perizia AI-assistita

Laboratorio

- Perizia simulata su output di AI, con vizi deliberati
- Un gruppo la difende (CTU), l'altro la attacca (CTP)

DELIVERABLE: memoria di parte (2 pagine)

Letture obbligatorie

- AI Act: estratti sui sistemi ad alto rischio
- Linee guida EDPB sul trattamento in ambito investigativo
- Giurisprudenza selezionata sulla prova scientifica (dispensa)

Verifica

Si valuta il rigore argomentativo di entrambi i gruppi, in attacco e in difesa.

Prossimo modulo: laboratorio integrato e casi di studio.

Un'analisi AI è difendibile in giudizio se i suoi errori sono misurati e dichiarati da un perito che ne risponde.



Grazie per l'attenzione

Vincenzo Calabrò

Domande, materiali del corso e contatti
sul sito del docente



vincenzocalabro.it