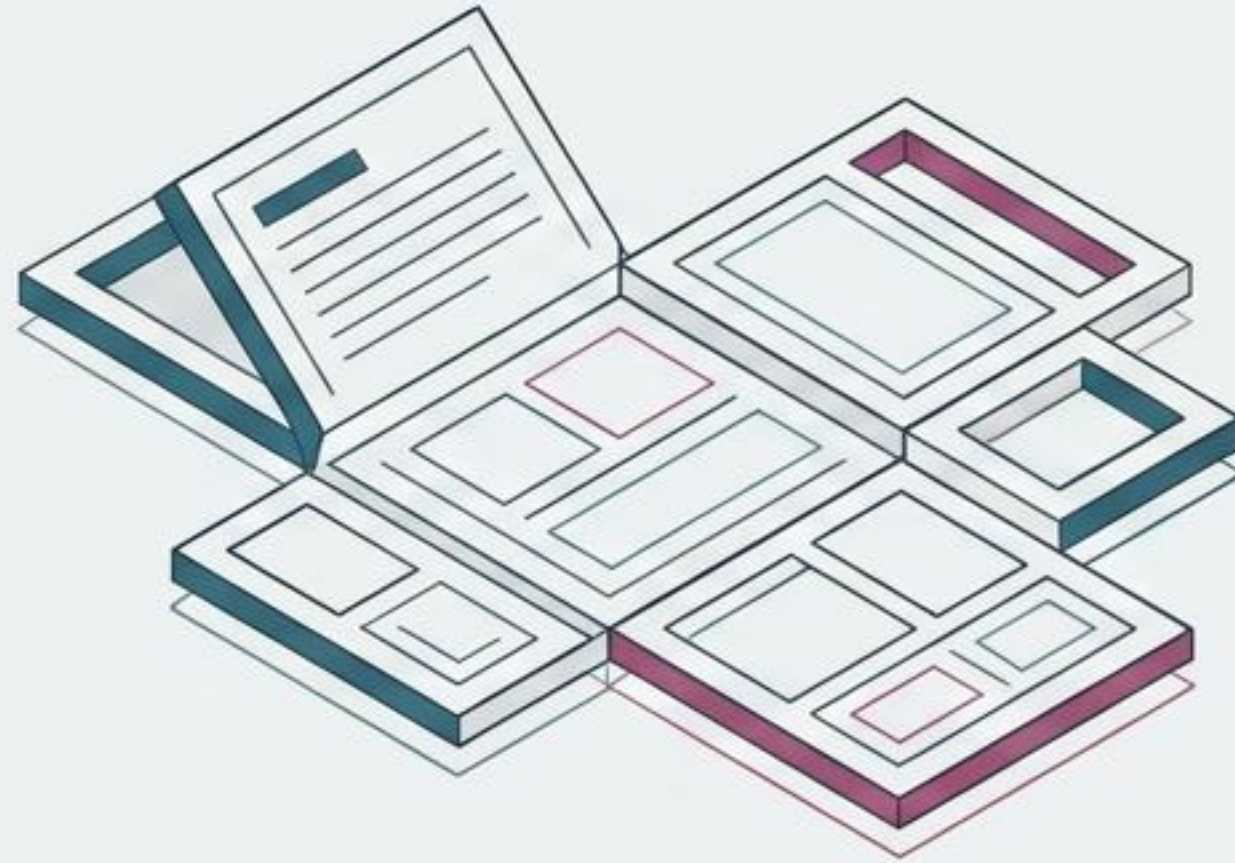


Il Blueprint della Trasparenza per l'IA

Guida procedurale al Codice di Condotta (Articolo 50 AI Act).

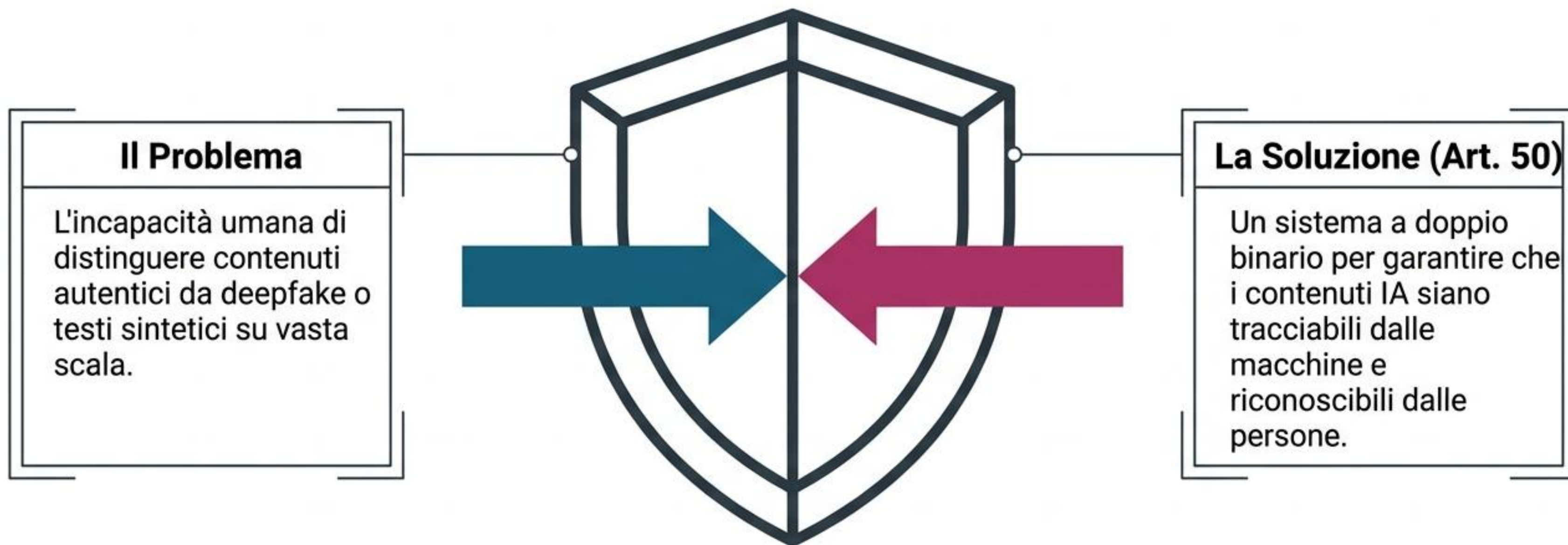


Un framework tecnico e visivo per l'implementazione delle normative europee su marcatura, rilevamento ed etichettatura dei contenuti sintetici.

Destinato a Fornitori di Sistemi IA e Soggetti che impiegano l'IA (Deployers)

Proteggere l'Ecosistema dell'Informazione

L'articolo 50 del Regolamento sull'IA stabilisce un mandato chiaro: proteggere la società dalla disinformazione. La trasparenza è l'architettura fondamentale per la fiducia digitale.



La Matrice delle Responsabilità (Art. 50)

Fornitori (Providers) - Sezione 1



Target

Sistemi di IA Generativa.



Obbligo

Marcatura leggibile da macchina (Art. 50.2).



Output

Metadati, Watermark invisibili, Strumenti di Rilevamento.



Natura

Invisibile e tecnica.

Soggetti (Deployers) - Sezione 2



Target

Deepfake e Testi di Interesse Pubblico.



Obbligo

Etichettatura visibile all'utente (Art. 50.4).



Output

Icona UE, Avvisi visivi e sonori.

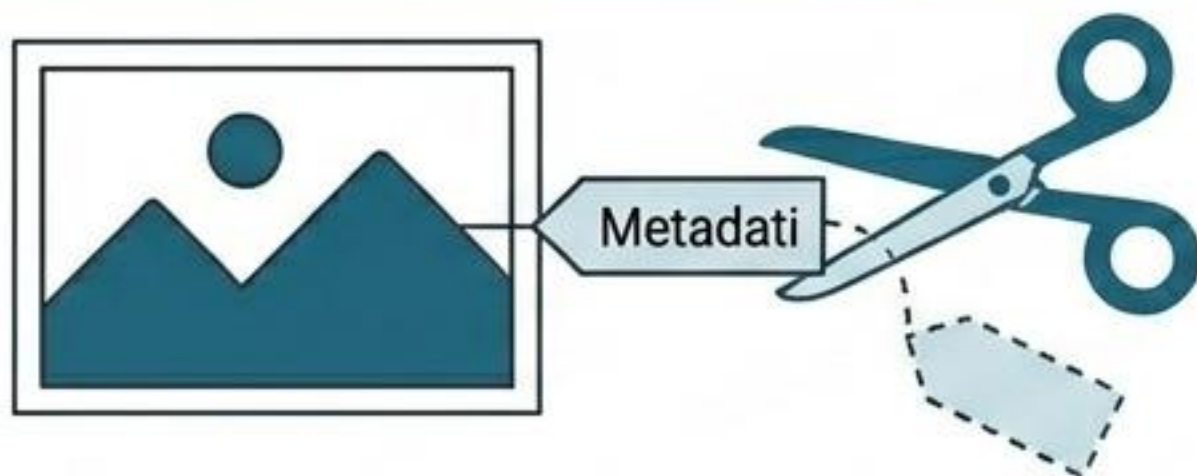


Natura

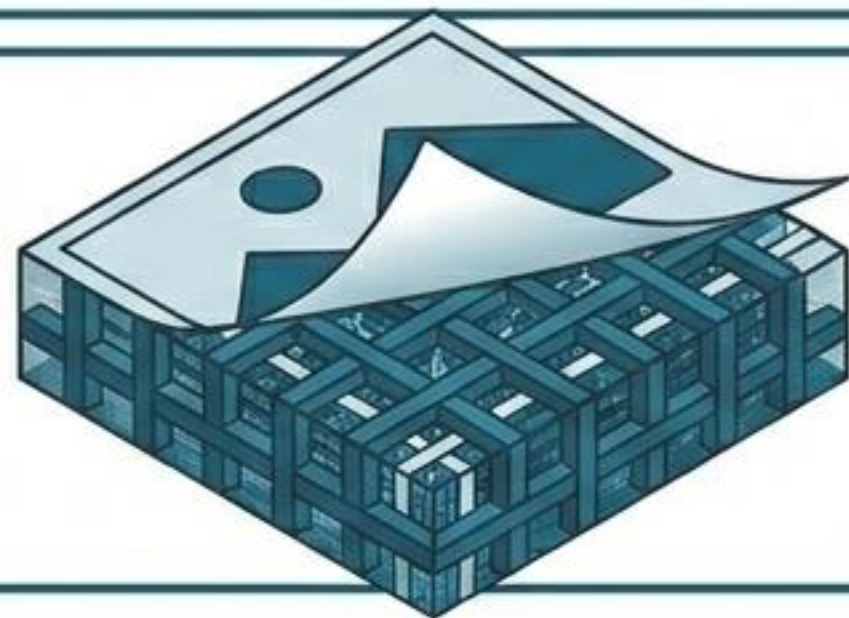
Visibile e umana.

Track 1: L'Approccio Multilivello (Fornitori)

Nessuna singola tecnica di marcatura può garantire da sola efficacia, interoperabilità e robustezza contro le manomissioni.



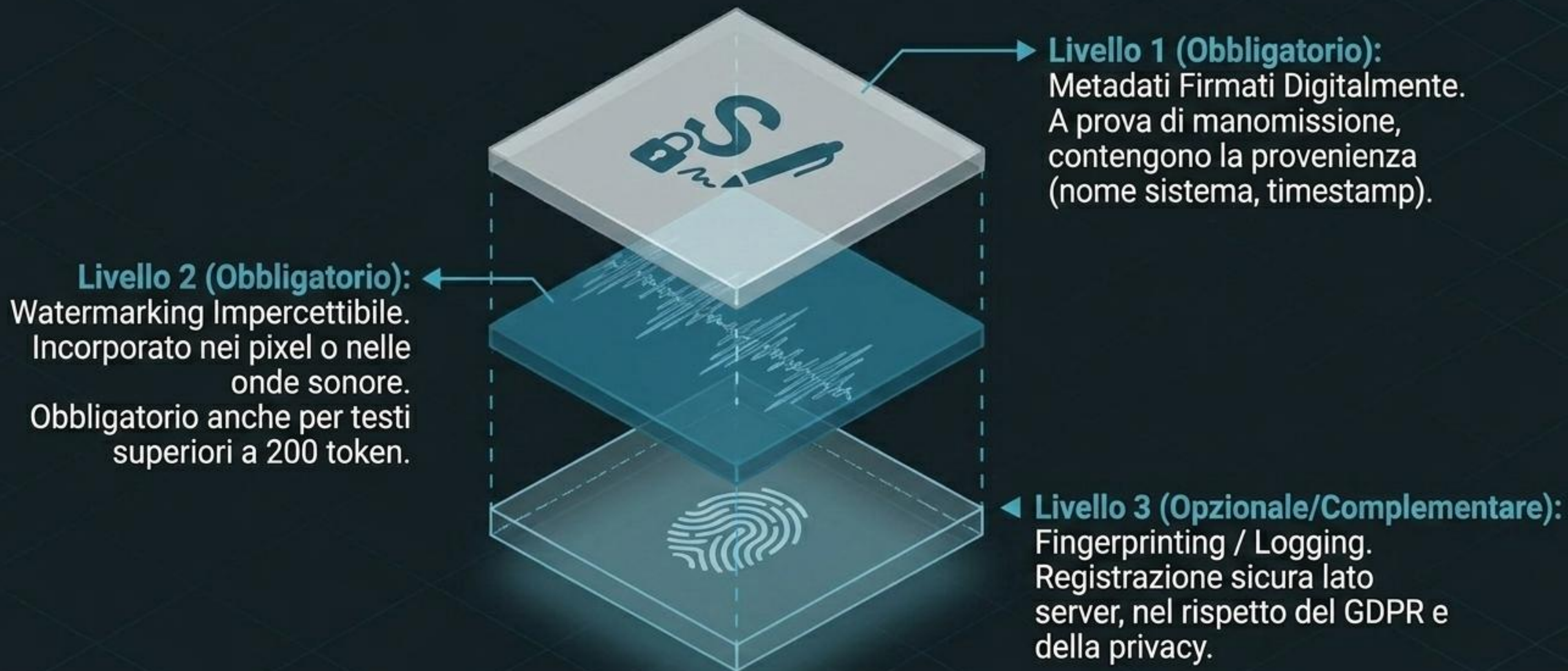
Singolo Livello: I metadati possono essere facilmente rimossi.



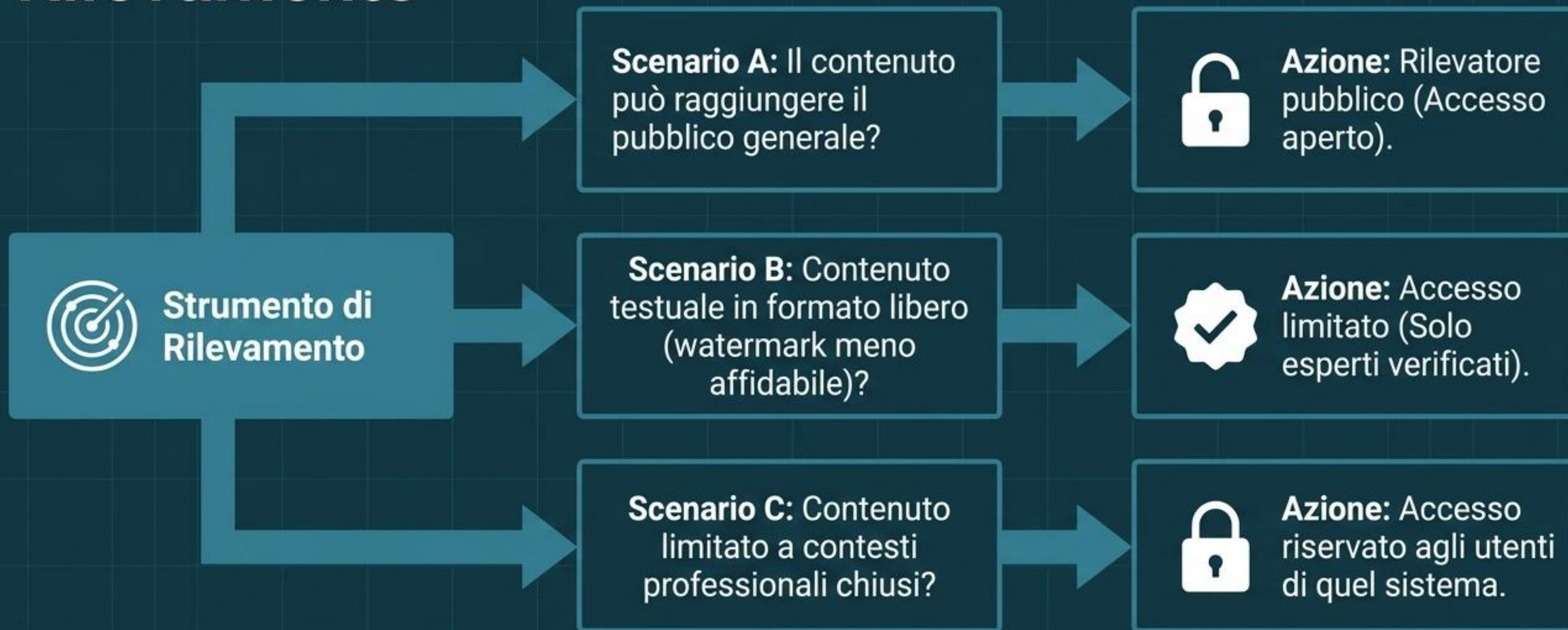
Multilivello: Il watermark impercettibile resiste alla manipolazione del file.

Obbligo: Implementare almeno due livelli di marcatura leggibile da macchina per i contenuti destinati alla diffusione online.

Lo Stack Tecnico di Marcatura



Accesso agli Strumenti di Rilevamento



Nota: È sempre obbligatorio garantire l'accesso alle Autorità di Sorveglianza del Mercato.

KPI per i Requisiti di Qualità



Efficacia

Misurata rigorosamente tramite il tasso di rilevamento riuscito (True Positive Rate) su contenuti alterati.



Affidabilità

Misurata tramite il tasso di errore (Falsi positivi e negativi) testato su campioni altamente diversificati.



Robustezza

Resistenza testata contro compressione, ritaglio, alterazioni comuni e attacchi avversari mirati.



Interoperabilità

Capacità tecnica di essere rilevato da sistemi terzi utilizzando protocolli e standard aperti.

Roadmap dell'Interoperabilità

Fase Iniziale

Adozione di standard aperti per i metadati.
Pubblicazione delle istruzioni di integrazione API.

Fase Intermedia

Sviluppo di segnali pubblici leggibili nei contenuti IA per instradare gli utenti al rilevatore corretto.

Febbraio 2027

Deadline: Implementazione obbligatoria di una soluzione condivisa.
Opzioni: Segnale interoperabile, Consorzio di rilevamento condiviso, o API aperta standardizzata.

Checklist di Conformità Interna (Fornitori)



1. Test Pre-Lancio

Verificare le soluzioni di marcatura contro i 4 KPI di qualità prima dell'immissione sul mercato.



2. Documentazione Architettuale

Mantenere un registro aggiornato che dimostri come le misure del Codice sono state tecnicamente implementate.



3. Formazione del Personale

Istruire i team di ingegneria sulle implicazioni dell'Articolo 50 e sulle best practice anti-elusione.

Track 2: L'Obbligo di Etichettatura Visibile (Soggetti)

Il ruolo del Deployer si attiva al momento della pubblicazione. La divulgazione deve avvenire al più tardi alla prima interazione dell'utente.



Trigger 1: Deepfake

Immagini, audio o video generati o manipolati dall'IA che sembrano falsamente autentici a una persona fisica.



Trigger 2: Testi Pubblicati

Testi su questioni di interesse pubblico generati senza revisione umana o controllo editoriale.

Specifiche di Design e Accessibilità



Accessibilità: Supporto obbligatorio per screen reader, descrizioni audio e segnali aptici.

Regole di Collocazione (Wireframes)



Video: Etichetta sovrapposta in un angolo per l'intera durata (se possibile). Non deve oscurare l'azione.

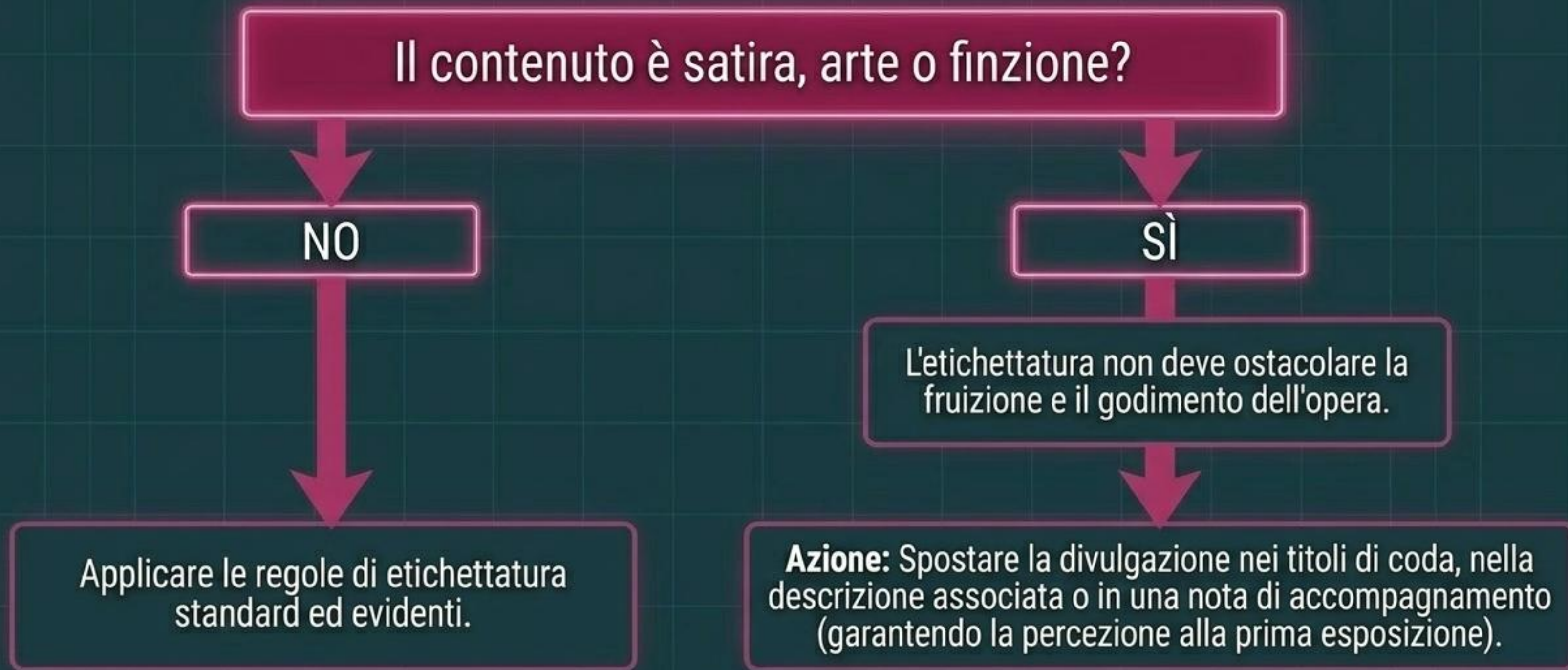


Immagini: Filigrana visibile agli angoli, fissa nel quadro visivo per evitare ritagli accidentali.



Testo Pubblicato: Collocazione nel colophon, nell'intestazione o immediatamente sotto il titolo principale.

Eccezioni e Opere Creative



Obbligo di Riesame: I Deployers devono istituire processi interni rigorosi per correggere tempestivamente i contenuti privi di etichetta o etichettati erroneamente.

L'Ecosistema della Trasparenza in Azione



Checklist Operativa Finale

Fornitori di IA Generativa

☐ Applica Metadati firmati digitalmente e Watermark impercettibile ai contenuti.

☐ Rendi pubblico e accessibile lo strumento di rilevamento.

☐ Verifica rigorosamente i KPI di Efficacia, Affidabilità e Robustezza.

☐ Pubblica e aggiorna la documentazione architetturale interna.

Soggetti / Deployers

☐ Integra l'Icona UE (o etichetta equivalente) in modo fisso nella UI.

☐ Applica avvisi sonori verbali per i contenuti esclusivamente audio.

☐ Adatta il posizionamento dell'etichetta per le opere satiriche e artistiche.

☐ Istituisce un processo interno di correzione rapida per le etichette errate.