Il Cibernetico

«Dimostra che l'hai scritto tu.»

Il rilevatore di IA non è una prova

Edoardo Ferri
Direttore Responsabile, Il Cibernetico · Consulente tecnico forense

E-PRIVACY XXXIX · PAVIA
ITIS G. CARDANO · 13 NOVEMBRE 2026

Lo stato dell'arte dei rilevatori

9%

OpenAI AI Classifier, 2023: testi umani segnalati come IA. Rilevava solo il 26% dei testi IA. Ritirato il 20 luglio 2023.

61,3%

Liang et al., Patterns 2023: falsi positivi medi di 7 rilevatori su saggi TOEFL di autori non madrelingua.

0,0041%

Pangram 4, luglio 2026: falsi positivi dichiarati dal produttore su 1 milione di testi web in inglese.

Il migliore oggi è molto accurato. Per un perito restano quattro domande.

Chi misura

Tassi del produttore, su testi web pre-2022: non su temi scolastici in italiano.

Quale lingua

Pangram 4.1: falsi negativi 0,37% in inglese, 1,99% su 18 lingue.

Quale versione

3.3 a maggio, 4 a luglio, 4.1 a ottobre 2026: tre modelli in sei mesi.

Che cosa produce

Probabilità per segmento: Human, AI-Assisted, AI-Generated.

Fonti: OpenAI (2023); Liang et al., Patterns (2023); Pangram Labs, model card Pangram 4 e 4.1 (2026)

Il punteggio come reperto: le garanzie che mancano

Principio ISO/IEC 27037	In una perizia	Con un AI detector in classe
Verificabilità (auditability)	Ogni operazione documentata: verbale, hash, catena di custodia	Testo incollato in una web app: nessun hash, nessun log, nessuna copia del reperto
Ripetibilità (repeatability)	Stesso operatore, stesso strumento, stesso risultato	Il fornitore aggiorna il modello; versione e soglia non vengono annotate
Riproducibilità (reproducibility)	Un altro tecnico, con strumenti equivalenti, ottiene un risultato confrontabile	Rilevatori diversi danno punteggi diversi; il modello è chiuso
Giustificabilità (justifiability)	Ogni scelta tecnica è motivata e difendibile	Chi decide non sa spiegare da dove viene il numero

ISO/IEC 27041 e 27042

Il metodo va validato prima dell'uso; l'interpretazione dichiara limiti, incertezza e tasso d'errore.

Contraddittorio · DPR 249/1998, art. 4, c. 3

Nessuna sanzione disciplinare senza aver prima invitato lo studente a esporre le proprie ragioni.

Fonti: ISO/IEC 27037:2012 (EN ISO/IEC 27037:2017), ISO/IEC 27041:2015, ISO/IEC 27042:2015; DPR 249/1998

La scuola come deployer: il quadro di riferimento

AI Act · Reg. (UE) 2024/1689

Allegato III, punto 3: valutare i risultati dell'apprendimento e rilevare comportamenti vietati durante le prove sono usi ad alto rischio. Obblighi dal 2 dicembre 2027 (Reg. (UE) 2026/1744).

Legge 132/2025, art. 14

Nella pubblica amministrazione l'IA è strumentale: conoscibilità del funzionamento e tracciabilità dell'uso. Decisione e responsabilità restano alla persona.

Linee guida MIM · DM 166/2025

La scuola è titolare del trattamento: DPIA e, per i sistemi ad alto rischio, integrazione con la FRIA (art. 27 AI Act).

ISO/IEC 42001:2023

Valutazione del rischio (6.1.2), valutazione d'impatto su persone e gruppi (6.1.4, metodo in ISO/IEC 42005), uso dei sistemi (A.9) e fornitori (A.10).

Già oggi la decisione resta del docente: motivata, tracciabile, conoscibile dallo studente.

Fonti: Reg. (UE) 2024/1689 e 2026/1744; L. 23 settembre 2025, n. 132; DM 9 agosto 2025, n. 166; ISO/IEC 42001:2023, 42005:2025

Analisi del rischio minima per una scuola pubblica

Rischio	Conseguenza	Misura minima
Falso positivo	Voto o sanzione ingiusti a un minore; danno reputazionale	Il punteggio non è mai l'unico elemento: colloquio sul testo e motivazione scritta
Bias su non madrelingua, DSA, scrittura semplice	Discriminazione indiretta	Valutazione d'impatto prima dell'adozione (ISO/IEC 42001, 6.1.4; FRIA)
Elaborati di minori su servizi esterni	Trattamento senza contratto; possibile trasferimento extra UE	Solo strumenti adottati dalla scuola, con nomina ex art. 28 GDPR e DPIA; nessun account personale
Versione e soglia non registrate	Esito non ripetibile né difendibile	Registro: strumento, versione, data, soglia, hash del testo; conservare l'originale

Il punteggio può aprire una verifica, non chiuderla. L'onere della prova resta a chi contesta.

Fonti: Reg. (UE) 2016/679, artt. 22, 28, 35; ISO/IEC 42001:2023; Linee guida MIM (DM 166/2025)

Grazie dell'attenzione

Edoardo Ferri

Direttore Responsabile, Il Cibernetico

Consulente tecnico forense in materia informatica · CTU presso il Tribunale di Pavia

EMAIL
mail@edoardoferri.it

RIVISTA
ilcibernetico.it

SITO PERSONALE
edoardoferri.it

LINKEDIN
linkedin.com/in/edoardoferri