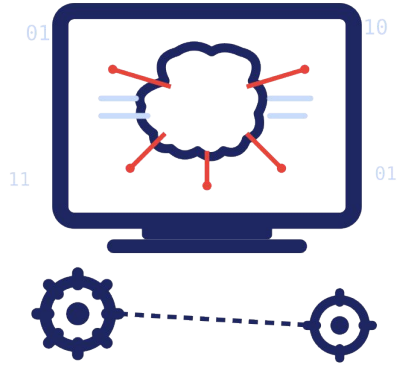


Dieci principi per un uso responsabile degli strumenti di rilevamento AI

GianMarco Schiesaro

Istat

Il divario tra promesse e realtà



Sistemi di rilevamento AI

Perplexity

prevedibilità del testo

Burstiness

irregolarità ritmica

98–99%

Accuratezza dichiarata

< 80%

Accuratezza reale



Vulnerabilità: la parafrasi

Riduce l'accuratezza di oltre il 20%.
Chi vuole ingannare il sistema può farlo

Watermarking (es. SynthID)

Filigrana digitale alla fonte. Promettente,
ma sensibile a parafrasi e traduzione

*La Dichiarazione di
Indipendenza americana viene
regolarmente classificata
come generata da AI*

! PRINCIPIO

Non trattare i punteggi come prove

I risultati sono indicatori probabilistici, non certezze. Nessun procedimento dovrebbe fondarsi esclusivamente sui punteggi

Chi paga il prezzo degli errori

Non madrelingua

Strutture semplici e vocabolario prevedibile scambiati per output AI

Minoranze etniche

Varianti linguistiche e dialetti non riconosciuti



Neurodivergenti

Stile metodico, letterale e ripetitivo interpretato come artificiale

Ricercatori STEM

Rigore scientifico e linearità scambiati per uniformità artificiale

PRINCIPIO

Proteggere le popolazioni vulnerabili

Soglie differenziate, revisione umana, procedure che tengano conto del contesto

PRINCIPIO

Testare su popolazioni diverse prima dell'adozione

Valutazione interna su un campione rappresentativo della propria popolazione

Sovranità digitale e protezione dei dati

Il problema

- Dati trasmessi a server esteri
- Riutilizzo per addestramento
- Conformità GDPR incerta
- Watermarking + blockchain = sorveglianza?



Le alternative aperte

- Binoculars – on-premise
- DetectGPT – nessun dato trasmesso
- SynthID – watermarking open source
- Verificabili, correggibili, adattabili

PRINCIPIO

Privilegiare strumenti aperti e verificabili

Limiti misurabili, comprensibili e correggibili. Sovranità sui dati

PRINCIPIO

Non basarsi mai su un singolo strumento

L'utilizzo aggregato di tre o più rilevatori riduce drasticamente il tasso di falsi positivi

Cambiare approccio



Il rilevamento AI ha limiti matematici intrinseci. La «corsa agli armamenti» tra software è insostenibile

Esami orali e in presenza



Argomentare e rispondere a domande impreviste

Valutazioni di processo



Evidenze raccolte lungo l'intero percorso

Portfolio



Evoluzione delle competenze nel tempo

Casi di studio e pratica



Applicare conoscenze a situazioni nuove



PRINCIPIO

Riprogettare le metodologie di valutazione

Le competenze nell'era AI sono diverse. Valutare il pensiero critico, non il prodotto testuale






PRINCIPIO

Educare all'uso responsabile dell'intelligenza artificiale




Superare la logica della minaccia da contrastare per abbracciare una visione educativa

Il framework completo: dieci principi

CAUTELA TECNOLOGICA

- 1  Non basarsi mai su un singolo strumento
- 2  Non trattare i punteggi come prove
- 3  Testare su popolazioni diverse prima dell'adozione

TUTELA DEI DIRITTI

- 4  Garantire il diritto di difesa e il contraddittorio
- 5  Proteggere le popolazioni vulnerabili
- 6  Garantire trasparenza e protezione dei dati

TRASFORMAZIONE

- 7  Privilegiare strumenti aperti e verificabili
- 8  Riprogettare le metodologie di valutazione
- 9  Formare docenti, ricercatori e valutatori
- 10  Educare all'uso responsabile dell'intelligenza artificiale



Paper completo



Dieci principi

Per le domande:
wooclap.com



Codice evento: **CONFARR26**