

Scheda tecnica per il responsabile della protezione dei dati

ilLasciapassare, anonimizzatore di documenti, versione 4.4

Versione 1.6 della scheda, 27 settembre 2026.

Questa scheda è scritta per chi, in un ente, deve valutare lo strumento prima di autorizzarne l'uso: il responsabile della protezione dei dati, il referente informatico, il responsabile qualità. Dice che cosa fa il programma, dove stanno i dati, quali reti tocca, come se ne verifica l'integrità e quali sono i risultati misurati, compresi i limiti. Per ogni affermazione è indicato, dove possibile, il modo di verificarla da soli.

Che cosa fa

Il programma copre i dati personali contenuti nei testi del lavoro educativo, scolastico e sociale (relazioni, verbali, piani educativi individualizzati, relazioni cliniche) prima che il testo venga affidato a uno strumento esterno di intelligenza artificiale, e li rimette al loro posto quando la risposta torna indietro. La sostituzione è reversibile soltanto sul computer dove avviene: ciò che esce è un testo con segnaposto al posto dei dati, mentre la corrispondenza fra segnaposto e dati veri resta in una tabella locale che non viene trasmessa a nessuno. In termini giuridici, verso l'esterno si tratta di una pseudonimizzazione ai sensi dell'articolo 4, numero 5, del Regolamento (UE) 2016/679, nella quale la chiave di ricostruzione non lascia mai la postazione; il nome «anonimizzatore» descrive l'uso corrente del termine nei servizi, non una qualificazione giuridica del risultato.

Il riconoscimento lavora su quattro livelli: un modello linguistico locale che copre ventidue categorie di dati (dai nomi di persona ai codici fiscali), regole deterministiche per le forme italiane che i modelli sbagliano (indirizzi, date, protocolli, IBAN e simili, con controllo matematico dove esiste), un glossario di termini che l'operatore insegna all'applicazione, e un elenco di esclusioni. Un quinto livello, separato, riguarda le categorie particolari degli articoli 9 e 10 del Regolamento (salute, vita sessuale, religione, opinioni, condanne, più la condizione economica e i dettagli che rendono riconoscibili): per ciascuno dei nove temi l'ente sceglie se ignorarlo, segnalarlo lasciandolo nel testo, o coprirlo come i nomi. L'impostazione di partenza è la segnalazione, perché in un documento il contenuto clinico è spesso il senso stesso del testo, e la scelta di farlo uscire deve restare una scelta umana e visibile.

Questa scelta è anche la differenza rispetto agli strumenti di anonimizzazione di uso generale, e conviene dirla con precisione, perché è la domanda che un ente pone per prima. Quegli strumenti riconoscono gli elementi che identificano una persona per la loro forma, cioè i nomi, i codici a struttura fissa, gli indirizzi e i recapiti, e li sostituiscono; alcuni dichiarano anche dizionari sanitari, che però contengono numeri e sigle, come la tessera o il referto. Nei documenti del lavoro educativo e sociale, invece, ciò che espone una persona è spesso una frase intera, e una frase non ha una forma che un riconoscitore di formati possa intercettare. I nove temi del quinto livello servono a portarla davanti a chi scrive prima che il testo esca, lasciando a lei la decisione.

Le due versioni

Il programma esiste in due versioni distinte, che si installano separatamente e non si mescolano, ciascuna con la propria cartella dati.

La **versione completa** è quella descritta in questa scheda, e non funziona senza licenza.

La **versione libera** si scarica liberamente, non scade e non ha licenza. Copre soltanto i nomi di persona in ogni loro forma, le iniziali puntate comprese, e le date. Tutto il resto di ciò che riconosce, cioè l'età, i luoghi (comune, via, civico, codice postale, provincia), i codici che identificano da soli (codice fiscale, IBAN, carta, documento, posta elettronica, telefono) e i nomi degli enti, insieme a tutti e nove i temi delle categorie particolari, lo **segnala e lo conta senza coprirlo**, lasciandolo nel testo. È una scelta dichiarata: una copertura parziale silenziosa produrrebbe testi che sembrano puliti e non lo sono, ed è il rischio peggiore per un ente. Chi valuta la versione libera deve quindi sapere che il testo che ne esce **non è pseudonimizzato per intero** e non va affidato a uno strumento esterno senza una rilettura umana. La versione libera non legge documenti, non conserva i lavori e non impara termini nuovi da coprire: ricorda soltanto i termini che chi la usa dichiara non personali, perché smetta di segnalarli.

Il confine fra le due versioni è fatto valere dentro il programma e non nella pagina che si vede: non lo si sposta né cambiando le impostazioni né insegnando termini, e il banco di prova lo verifica a ogni costruzione.

Dove stanno i dati

Tutto sul computer dell'operatore: il programma, il modello di riconoscimento (che sta sul disco e non viene scaricato all'avvio), i testi lavorati, le tabelle di ripristino, il glossario, i lavori salvati. Il programma ascolta soltanto sull'indirizzo locale della macchina (127.0.0.1), non accetta richieste dall'esterno, respinge le richieste che non portano un'intestazione Host locale (contro il DNS rebinding) e protegge ogni chiamata con un gettone di sessione

che solo la propria pagina conosce (contro le richieste da siti estranei aperti nello stesso browser); la pagina porta inoltre le intestazioni di sicurezza che impediscono di incorniciarla in un altro sito. Il programma non invia dati in rete: né i testi, né statistiche d'uso, né segnalazioni di errore, e non contiene telemetria; il canale di diagnostica che la libreria di esecuzione del modello prevede sulle versioni per Windows viene spento esplicitamente all'avvio. Il componente di Windows che disegna la finestra del programma ha collegamenti propri, descritti alla voce «Quali reti tocca».

Il programma non scrive nulla fuori dalla propria cartella dati. Non tocca il registro di Windows, non lascia file nella cartella personale né nelle cartelle comuni del sistema, e disinstallandolo non resta traccia oltre alla cartella dati, che l'utente può cancellare quando vuole. Fino alla versione 4.0 il programma annotava il primo utilizzo in sei sedi di sistema, perché la prova gratuita di sette giorni non ricominciasse disinstallando e riscaricando: la prova non esiste più, e con essa quelle annotazioni. Chi avesse installato una versione fino alla 4.0 può togliere le annotazioni rimaste con lo strumento «ripulisci-segni-vecchi.py», che il licenziante fornisce su richiesta.

Dentro la cartella dati il programma annota il giorno di attivazione della licenza, l'ultimo giorno in cui ha lavorato e il codice della postazione (una sigla derivata dall'identificatore della macchina, non riferita a una persona): annota il giorno in cui ciascuna licenza è stata attivata, da cui si conta la durata pattuita, e l'ultimo giorno in cui ha lavorato. Restano sul computer e non vengono trasmessi a nessuno. Che il programma non scriva fuori da lì non è una dichiarazione ma un fatto verificato: il banco di prova lo controlla a ogni costruzione, cercando quelle sei sedi e pretendendo che nessuna esista.

La tabella di ripristino della sessione in corso sta nella memoria della pagina e viene cancellata quando la pagina si chiude o si ripulisce; i lavori che l'operatore decide di salvare stanno nella cartella dati, come il glossario. Quei file contengono i nomi veri in chiaro e nascono leggibili soltanto dall'utente che li ha scritti (permessi del profilo su Windows e Mac): la loro protezione a riposo si appoggia alle misure della postazione, in primo luogo la cifratura del disco (BitLocker, FileVault, LUKS), che si raccomanda. L'interfaccia si apre in una finestra propria del programma, disegnata dal componente del sistema operativo; soltanto se quel componente manca il programma ripiega sul browser dell'operatore, e in quel caso, poiché le estensioni del browser leggono le pagine anche su 127.0.0.1, si raccomanda un profilo senza estensioni. Le aree di testo disattivano il correttore ortografico e il completamento automatico, che altrimenti potrebbero inviare il contenuto ai server del fornitore.

Quali reti tocca

Il programma non apre collegamenti verso l'esterno: ascolta soltanto sull'indirizzo locale del computer, non si collega a internet per funzionare, per caricare il modello né per verificare la licenza, che è un file firmato crittograficamente (Ed25519) e viene controllato in locale, e non trasmette a nessuno i testi, le tabelle di ripristino o i termini imparati. Si può staccare la rete e usarlo dall'inizio alla fine, copertura e ripristino compresi.

Su Windows la finestra del programma è disegnata da Microsoft Edge WebView2, un componente del sistema operativo, e chi osserva il traffico di rete vede partire da quel componente, a ogni avvio, alcuni collegamenti verso servizi di Microsoft: la configurazione e l'aggiornamento del componente stesso e, quando sul computer è attivo un account Microsoft, i servizi di quell'account, con la risoluzione dei nomi che passa dal DNS impostato sul computer. Il componente li apre per conto proprio, secondo quanto Microsoft descrive nella pagina «Data and privacy in WebView2», e una parte di quei dati, dice Microsoft, viene raccolta qualunque sia l'impostazione di diagnostica di Windows. Il programma spegne le due funzioni che Microsoft lascia governare a chi scrive il programma: il controllo SmartScreen delle pagine, inutile per una pagina servita dallo stesso computer, e l'invio a Microsoft dei rapporti di errore del componente, che restano sul computer. Su macOS la finestra è disegnata dal componente WebKit del sistema.

Come si verifica l'integrità di ciò che si riceve

A ogni rilascio vengono pubblicate le impronte SHA-256 dei pacchetti, con una pagina di istruzioni che spiega come ricalcolarle su Windows e Mac senza competenze tecniche. Tutti i pacchetti vengono costruiti dalle macchine di GitHub, non sul computer dell'autore, a partire dal codice depositato, e ogni costruzione esegue prima di consegnare il pacchetto un collaudo di oltre cinquanta prove (ingressi ostili, documenti malformati, sicurezza, licenze, carico), fra le quali il controllo che il programma non scriva più nelle sei sedi di sistema in cui le versioni fino alla 4.0 annotavano il primo utilizzo; il registro di ogni costruzione è consultabile, e su richiesta il registro e il codice sorgente possono essere mostrati ai tecnici dell'ente. Il modello di riconoscimento, convertito una volta sul computer dell'autore, è custodito nel deposito con la propria impronta. Su Windows il programma si installa dal Microsoft Store, che verifica e firma il pacchetto prima di distribuirlo; il pacchetto per macOS è firmato con il certificato di sviluppatore dell'autore e notarizzato da Apple. Per i pacchetti scaricati dal sito le impronte pubblicate consentono la verifica indipendente di ciò che si è ricevuto.

I risultati misurati, e i limiti

La qualità del riconoscimento è misurata su un banco di prova di cinquanta documenti fittizi delle quattro tipologie ricorrenti in questi ambiti (relazioni di supervisione, verbali d'équipe, relazioni per l'autorità giudiziaria, progetti educativi individualizzati), con 864 dati personali annotati e casi difficili inseriti apposta: nomi ripetuti da soli, nomi non italiani, sigle ambigue, cognomi che sono anche nomi di città, nomi composti, IBAN a gruppi, date per esteso, numeriche e senza anno. Il criterio è il più severo possibile: un dato conta come sfuggito se anche un solo suo frammento sopravvive nel testo coperto.

Sul pacchetto 4.4, quello descritto da questa scheda, la misura dà un richiamo complessivo del 98,5 per cento (13 dati sfuggiti su 864, con il glossario vuoto, cioè senza i termini che l'operatore insegna con l'uso) e una sola sostituzione non necessaria su 901 sostituzioni. I dati sfuggiti si concentrano nei nomi di enti e servizi, la categoria più ambigua.

Questi numeri vanno letti per quello che sono: nessun sistema automatico, di alcun produttore, riconosce ogni dato in ogni testo. Per questo il programma mostra al termine di ogni analisi un pannello «Da rileggere» con ciò che merita un occhio umano, e la revisione finale resta in capo all'operatore. Va detto con la stessa franchezza che un testo può rendere riconoscibile una persona anche senza identificativi diretti, attraverso il contenuto: è la ragione del quinto livello di segnalazione, e insieme il motivo per cui lo strumento riduce il rischio ma non sostituisce la valutazione dell'ente su che cosa affidare a un fornitore esterno.

Licenze, costi e continuità del servizio

Il programma è concesso in licenza d'uso annuale, non venduto. Le formule sono quattro: 75 euro l'anno per una postazione singola, 340 euro per un ente fino a cinque postazioni, 930 fino a quindici, 2.360 fino a quaranta; oltre quel numero, e per la formula che unisce un percorso formativo alla licenza d'ente, si procede a preventivo. Gli importi sono finali, senza IVA, perché il licenziante opera in regime forfettario, e sono comprensivi degli aggiornamenti del programma, delle regole e dei dizionari e dell'assistenza per tutta la durata della licenza. Il numero di postazioni pattuito è indicato nell'ordine e vincola l'uso, indipendentemente dalla forma del file di licenza.

Due elementi rilevanti per la valutazione di continuità. Il primo: la licenza vale per il tempo dichiarato e **alla scadenza il programma si ferma per intero**, cessando sia la copertura dei dati sia il loro ripristino; l'ente deve quindi organizzare il lavoro in modo che i testi affidati a strumenti esterni vengano ricomposti entro la validità della licenza, e il programma avvisa in modo visibile nei giorni che precedono la scadenza. Il secondo: i dati

dell'ente non vengono né cancellati né resi illeggibili, perché i termini insegnati, le impostazioni e i lavori salvati restano nelle cartelle dati delle postazioni, in formato leggibile, e tornano utilizzabili appena una licenza nuova viene inserita. Non esistono quindi né un formato proprietario che trattenga i dati né una dipendenza da servizi in rete che possano cessare.

Le correzioni di sicurezza restano gratuite per cinque anni dal rilascio di ciascuna versione, anche a licenza cessata, come dichiarato nelle condizioni.

L'inquadramento per il registro dei trattamenti

Lo strumento attua i principi di minimizzazione dei dati e di protezione fin dalla progettazione e per impostazione predefinita (articoli 5 e 25 del Regolamento) e può essere censito fra le misure tecniche e organizzative dell'articolo 32 per i trattamenti che prevedono l'uso di strumenti di intelligenza artificiale esterni. Il fornitore del programma non riceve, non vede e non tratta alcun dato dei testi lavorati: non è quindi un responsabile del trattamento, e non serve alcun contratto di nomina ai sensi dell'articolo 28. La titolarità resta interamente all'ente, compresa la scelta e la valutazione del fornitore di intelligenza artificiale verso cui i testi coperti vengono inviati; su quel fornitore, e non su questo programma, ricade la verifica che esista un contratto ai sensi dell'articolo 28, che gli abbonamenti destinati alle organizzazioni prevedono e che invece non accompagna gli accessi gratuiti destinati al consumatore.

Quanto al regolamento europeo sull'intelligenza artificiale (regolamento (UE) 2024/1689, modificato dal regolamento (UE) 2026/1744, il cosiddetto omnibus digitale, in vigore dal 27 luglio 2026), un riconoscitore locale di dati personali non rientra fra i sistemi ad alto rischio né fra quelli soggetti agli obblighi specifici di trasparenza dell'articolo 50, che si applicano dal 2 agosto 2026. Restano in capo all'ente due doveri già esigibili e uno che arriva. I due esigibili sono il dovere di alfabetizzazione del personale che usa sistemi di intelligenza artificiale, che l'articolo 4 impone dal 2 febbraio 2025 e che l'omnibus ha riformulato come obbligo di misure efficaci e proporzionate, calibrate sui ruoli e sulle conoscenze tecniche di chi opera, ed è la ragione per cui lo strumento viene offerto anche insieme a un percorso formativo; e il divieto, posto dall'articolo 5, di dedurre le emozioni delle persone nei luoghi di lavoro e negli istituti di istruzione, che nessuna funzione di questo programma esercita. Quello che arriva riguarda invece i sistemi ad alto rischio dell'allegato III, i cui obblighi l'omnibus ha spostato dal 2 agosto 2026 al 2 dicembre 2027: da quella data gli enti che impiegano sistemi di intelligenza artificiale per l'istruzione e la formazione professionale (allegato III, punto 3), oppure per valutare l'ammissibilità delle persone alle prestazioni e ai servizi di assistenza pubblica essenziali anche quando operano per conto di un'autorità

pubblica (punto 5, lettera a), risponderanno degli obblighi dell'articolo 26, e gli organismi di diritto pubblico insieme ai soggetti privati che forniscono servizi pubblici dovranno svolgere la valutazione d'impatto sui diritti fondamentali prevista dall'articolo 27. Questo programma non è uno di quei sistemi, ma il calendario serve a chi deve programmare gli adempimenti dell'ente.

Sul versante italiano, la legge 23 settembre 2025, n. 132 (Gazzetta Ufficiale n. 223 del 25 settembre 2025, in vigore dal 10 ottobre 2025), che il suo articolo 1 vuole interpretata e applicata conformemente al regolamento (UE) 2024/1689, aggiunge tre obblighi che riguardano l'ente e non il programma. Il datore di lavoro o il committente è tenuto a informare il lavoratore dell'utilizzo dell'intelligenza artificiale, nei casi e con le modalità dell'articolo 1-bis del decreto legislativo 26 maggio 1997, n. 152 (articolo 11, comma 2). Nelle professioni intellettuali le informazioni sui sistemi impiegati vanno comunicate al soggetto destinatario della prestazione con linguaggio chiaro, semplice ed esaustivo, e l'utilizzo è finalizzato al solo esercizio delle attività strumentali e di supporto, con prevalenza del lavoro intellettuale oggetto della prestazione d'opera (articolo 13); in ambito sanitario l'interessato ha diritto di essere informato sull'impiego di queste tecnologie (articolo 7, comma 3). Nella pubblica amministrazione, infine, l'intelligenza artificiale opera in funzione strumentale e di supporto all'attività provvedimentale, e la persona resta l'unica responsabile dei provvedimenti e dei procedimenti in cui sia stata utilizzata (articolo 14, comma 2). Nessuno dei tre obblighi grava su questo programma, che non invia i testi a nessuno e non decide nulla al posto di chi scrive; i primi due, tuttavia, si soddisfano più facilmente se l'ente sa dire quali strumenti impiega e a che titolo, ed è fra le ragioni per cui questa scheda esiste.

Sicurezza del programma

Il programma non richiede privilegi di amministratore e non scrive nulla fuori dalla cartella dati dell'utente. La lettura dei documenti Word e OpenDocument passa da una libreria che neutralizza i costrutti XML pericolosi (entità esterne, espansioni a valanga) e da un tetto alla dimensione decompressa, contro gli archivi costruiti ad arte; i testi incollati hanno un tetto dichiarato. Le librerie incorporate sono fissate per versione a ogni costruzione e verificate contro gli archivi pubblici delle vulnerabilità note; le licenze di ogni componente accompagnano il pacchetto, insieme alla distinta dei componenti. Eventuali vulnerabilità possono essere segnalate agli indirizzi indicati sotto, e vengono gestite secondo la procedura di segnalazione del fornitore: correzioni di sicurezza distribuite appena disponibili, con avviso per posta elettronica agli intestatari e impronte nuove, per un periodo di supporto di cinque anni dal rilascio di ciascuna versione, in linea con il

regolamento europeo sulla ciberresilienza (regolamento (UE) 2024/2847), al quale il prodotto si conforma per autovalutazione come prodotto della categoria ordinaria.

Contatti

Davide Cadeddu, consulente pedagogico, con domicilio professionale in Borgata Rocco 22, 10040 Rubiana (TO). Partita IVA 11961250013. Telefono +39 328 8435740. Posta elettronica: info@illasciapassare.it. Posta certificata: cadeddu.davide@pec.it. Su richiesta: pagina delle impronte e delle verifiche, registro di costruzione, codice sorgente in visione per i tecnici dell'ente, corpus di prova per una dimostrazione dal vivo su dati fittizi.