RELAZIONE DI CONFORMITÀ GDPR

Progetto EduBot

Analisi di conformità al Regolamento UE 2016/679

Titolare del Trattamento	Istituto Tecnico Industriale 'G. Omar' - Novara
Responsabile del Progetto	Prof. Edoardo Salza
Versione Software	EduBot v7.0 (VertexOnly)
Versione Documento	2.0
Data Ultima Revisione	Febbraio 2026
Classificazione	Uso Interno — ITI G. Omar

1. Architettura 'Stateless' e Limitazione della Conservazione

L'applicazione implementa una gestione della sessione puramente volatile attraverso la funzione st.session_state di Streamlit.

Caratteristiche tecniche

- Assenza di Database Persistente: Non sono presenti integrazioni con database (Firestore, SQL, ecc.). Tutte le informazioni (cronologia chat, file elaborati, token di autenticazione) risiedono esclusivamente nella memoria RAM del server durante la sessione attiva.
- Timeout Automatico: La classe Config definisce SESSION_TIMEOUT = 3600 (60 minuti). La funzione check_session_limits() invalida la sessione dopo tale periodo di inattività, forzando la cancellazione dei dati.
- Cancellazione al Logout: La funzione reset_session() garantisce la rimozione integrale di ogni chiave dallo stato di sessione, assicurando che nessun residuo di dati personali permanga dopo la chiusura del servizio.
- Gestione File Volatile: I file caricati sono elaborati in cache temporanea (OptimizedFileCache, max 150 MB, max 5 file) e cancellati automaticamente a fine sessione.

Conformità: Art. 5, c. 1, lett. e) GDPR — Limitazione della conservazione.

2. Minimizzazione dei Dati (Art. 5, c. 1, lett. c)

Il software è progettato per raccogliere solo i dati strettamente necessari all'erogazione del servizio.

Dati NON trattati

NO	Email dello studente
NO	Nome e cognome
NO	Dati anagrafici
NO	Indirizzo IP

Dati trattati (minimi)

SI	**Hash anonimo:** Identificativo alfanumerico pseudonimo generato dal token di accesso OAuth (es. 'a3f8d29b'), non reversibile e non ricollegabile all'identità dello studente.
SI	**Dominio istituzionale:** 'itiomar.net' (informazione generica, non identifica individui).
SI	**Session ID:** UUID casuale (uuid.uuid4().hex[:12]) per gestione tecnica della sessione.

Implementazione OAuth Privacy-First

Il processo di autenticazione è progettato per non estrarre dati personali:

- Lo studente clicca 'Accedi con Google'
- Google verifica che l'account sia @itiomar.net (restrizione di dominio configurata sul progetto OAuth — accesso riservato agli account istituzionali)
- Se autorizzato, Google emette un token di accesso (valido 60 minuti)
- L'applicazione genera un hash anonimo monodirezionale dalla firma del token
- Nessun dato personale (email, nome) viene estratto, salvato o processato

Anonimizzazione Automatica

Il sistema SecuritySystem.anonymize_data() rileva e oscura automaticamente eventuali dati sensibili (email, chiavi API) inseriti per errore dall'utente nei prompt prima che vengano inviati al modello AI.

3. Privacy by Design e by Default (Art. 25 GDPR)

Onboarding Bloccante

Il software impedisce l'accesso alla chat finché l'utente non ha esplicitamente accettato:

- Informativa completa sul trattamento dei dati personali (art. 13 GDPR)
- Condizioni d'uso e limitazioni del sistema AI
- Disclaimer sulla responsabilità delle risposte generate

Controllo degli Accessi

- Restrizione dominio OAuth: Solo account @itiomar.net autorizzati (configurazione a livello di progetto Google Cloud)
- Service Account con permessi minimi: Principio del privilegio minimo (least privilege)
- Nessuna gestione password: Autenticazione delegata a Google Identity Platform

Configurazione Predefinita Sicura

- Timeout sessione automatico (60 minuti)
- Cache file con limite rigido (150 MB, max 5 file)
- USE_VERTEX_AI hardcoded come costante nel codice — non sovrascrivibile tramite variabili d'ambiente

4. Integrità e Riservatezza (Art. 5, c. 1, lett. f)

A. Sicurezza delle Comunicazioni

- TLS 1.3: Crittografia end-to-end per tutte le comunicazioni
- HTTPS obbligatorio: Certificati gestiti automaticamente da Cloud Run

B. Sicurezza del Prompt

Il sistema implementa funzioni di sicurezza (SecuritySystem.detect_injection_with_ai()) per prevenire:

- Prompt injection (tentativi di manipolare le istruzioni di sistema)
- Estrazione del system prompt
- Utilizzi impropri del servizio

C. Gestione File Sicura

- I file caricati vengono processati in cache volatile con politica LRU (Least Recently Used), assicurando rimozione automatica dei file meno recenti
- PDF con formule matematiche: Convertiti in immagini ad alta risoluzione (144 DPI) per preservare le equazioni. Il modello AI è multimodale e 'vede' le formule nelle immagini.
- Nessun caricamento su server persistente: I file rimangono esclusivamente in memoria RAM
- Cancellazione automatica alla chiusura della sessione, timeout o tramite pulsante 'Reset Chat'

D. Pseudonimizzazione

Le sessioni sono identificate da hash casuali, rendendo impossibile:

- Correlare sessioni dello stesso utente
- Risalire all'identità reale dal session ID
- Tracciare comportamenti individuali nel tempo

5. Elaborazione AI tramite Google Cloud Vertex AI

L'applicazione utilizza Google Cloud Vertex AI (data center europe-west1 — Belgio) per l'elaborazione delle richieste AI, garantendo protezioni contrattuali e sovranità dei dati UE.

Il codice utilizza Vertex AI in modalità esclusiva: USE_VERTEX_AI è una costante hardcoded. In caso di indisponibilità del servizio, l'accesso viene bloccato per garantire la continuità delle protezioni contrattuali GDPR. Non è previsto alcun fallback verso servizi AI privi di Data Processing Addendum.

Protezioni Contrattuali (CDPA — Cloud Data Processing Addendum)

SI	Google agisce come Responsabile del Trattamento (art. 28 GDPR): non decide come usare i dati, esegue solo istruzioni del Titolare
SI	Google NON usa i dati per addestrare AI: clausola contrattuale vincolante (protezione enterprise, superiore ai servizi consumer)
SI	Sovranità dei dati UE: tutti i dati rimangono nell'Unione Europea (regione europe-west1, Belgio)
SI	Standard Contractual Clauses (SCC): trasferimenti extra-UE disciplinati da clausole approvate dalla Commissione Europea + crittografia end-to-end
SI	In caso di indisponibilità Vertex AI: accesso bloccato — nessun fallback verso servizi privi di DPA

6. Logging e Audit Trail

Per finalità di sicurezza e audit tecnico, il sistema genera log strutturati (log_critical_event()) che contengono esclusivamente dati anonimi minimali. Il logging è attivo solo in ambiente di produzione (Cloud Run); in locale non vengono generati log.

I 3 eventi loggati (esclusivamente questi)

Evento	Descrizione
authentication_success	Audit degli accessi al sistema (obbligatorio per conformità)
session_terminated	Prova di cancellazione dati a fine sessione (obbligatorio GDPR Art. 17)
critical_error	Solo crash o errori gravi del sistema (es. timeout API, errori 500)

Dati presenti nei log

SI	Timestamp: data e ora dell'evento
SI	Session hash anonimo: primi 6 caratteri dell'hash di sessione (non persistente, non ricollegabile)
SI	Dominio istituzionale: 'itiomar.net' (informazione generica)
SI	Tipo di evento: uno dei 3 eventi definiti sopra
NO	Contenuto delle conversazioni
NO	Email o nome utente
NO	Indirizzi IP
NO	File caricati o loro contenuto

Periodo di conservazione	7 giorni (bucket dedicato con retention automatica)
Bucket Sistemi Automatici	edubot-sistemi-audit-7days (europe-west1)
Bucket Laboratorio Robotica	edubot-labrob-audit-7days (europe-west1)
Base giuridica	Consenso (art. 6, c. 1, lett. a) + interesse pubblico (lett. e) GDPR

Filtro sink Cloud Logging	severity=WARNING AND textPayload=~"AUDIT:" AND resource.labels.service_name=[servizio]

7. Conformità AI Act (Reg. UE 2024/1689)

Classificazione	Sistema AI a rischio LIMITATO
Motivazione	Sistema che interagisce direttamente con utenti umani — obbligo di trasparenza (art. 50 AI Act)

Nota: la classificazione 'rischio limitato' (non 'minimo') è corretta per sistemi di interazione diretta con utenti. I sistemi a rischio minimo (es. filtri antispam) non hanno obblighi di trasparenza attivi. EduBot prevede invece disclosure esplicita dell'interazione con AI, rientrando nella categoria rischio limitato con obbligo di trasparenza ex art. 50.

Cosa NON fa il sistema

NO	Non valuta risultati di apprendimento
NO	Non determina accesso a percorsi formativi
NO	Non monitora studenti durante esami
NO	Non assegna voti o valutazioni formali

Trasparenza algoritmica (Art. 50 AI Act)

SI	L'utente è sempre informato di interagire con un sistema AI (nome 'EduBot', avatar robotico, disclaimer esplicito)
SI	Disclaimer permanente: 'Le risposte possono contenere errori — verifica sempre con fonti ufficiali'
SI	Human-in-the-loop: il docente mantiene la supervisione pedagogica completa
SI	Non influisce su decisioni che producono effetti giuridici (art. 22 GDPR)

Conformità: Linee Guida MIM (Ministero dell'Istruzione e del Merito, gennaio 2025) sull'utilizzo dell'AI nella scuola.

8. Diritti dell'Interessato (Artt. 15-22 GDPR)

Durante la sessione

SI	Accesso (art. 15): visualizzazione in tempo reale della cronologia conversazione
SI	Rettifica (art. 16): correzione tramite nuovi messaggi nella conversazione
SI	Esportazione: download conversazione in formato HTML con rendering formule matematiche (KaTeX)

In qualsiasi momento

SI	Cancellazione (art. 17): pulsante 'Reset Chat' — cancella cronologia e file mantenendo la sessione

SI	Cancellazione (art. 17): pulsante 'Logout' — termina sessione e cancella tutti i dati
SI	Cancellazione (art. 17): chiusura del browser — medesimo effetto del logout
SI	Diritto all'oblio: garantito automaticamente dall'architettura stateless (nessun dato persiste oltre la sessione)

Per dati OAuth gestiti da Google

L'utente esercita i propri diritti tramite le impostazioni dell'account Google (https://myaccount.google.com) → Account Google → Sicurezza → App di terze parti con accesso all'account.

Riepilogo di Conformità

GDPR (Reg. UE 2016/679)

SI	Accesso sicuro: Google OAuth 2.0 con restrizione di dominio @itiomar.net
SI	Informativa completa: visualizzata obbligatoriamente prima dell'uso con accettazione esplicita (art. 13)
SI	Minimizzazione dei dati: solo hash anonimo, dominio, session ID
SI	Nessun storage persistente: architettura stateless, dati solo in RAM
SI	Sessione temporanea: timeout 60 minuti, cancellazione automatica
SI	Sovranità dei dati: elaborazione AI in UE (Vertex AI — Belgio, europe-west1)
SI	Trasparenza: informativa dettagliata su finalità, base giuridica, destinatari
SI	Audit trail anonimo: 7 giorni retention, solo 3 eventi critici, nessun dato identificativo

AI Act (Reg. UE 2024/1689)

SI	Classificazione corretta: sistema a rischio limitato (interazione diretta con utenti)
SI	Trasparenza: utente sempre informato di interagire con AI
SI	Human oversight: docente mantiene supervisione pedagogica
SI	Nessuna decisione automatizzata: non influisce su valutazioni o accessi formativi

Best Practice di Sicurezza

SI	Crittografia: TLS 1.3, HTTPS obbligatorio
SI	Autenticazione forte: OAuth 2.0 delegata a Google Identity Platform
SI	Controllo accessi: Service Account con permessi minimi (least privilege)
SI	Audit trail: log anonimizzati (7 giorni, solo 3 eventi)
SI	Backend esclusivo: Vertex AI Enterprise con DPA — nessun fallback a servizi consumer

Servizi Google Utilizzati

Servizio	Funzione
Cloud Run	Hosting applicazione EduBot (Container as a Service)
Vertex AI	Elaborazione intelligenza artificiale (CDPA attivo, europe-west1)
Secret Manager	Gestione sicura credenziali OAuth
Cloud Logging	Audit trail sicurezza (bucket 7 giorni)
Identity Platform	Autenticazione OAuth 2.0 con restrizione dominio

Certificazioni di Sicurezza e Privacy (Google Cloud)

Certificazione	Riferimento
ISO/IEC 27001:2013 — Information Security Management	https://cloud.google.com/security/compliance/iso-27001
ISO/IEC 27017:2015 — Cloud Security Controls	https://cloud.google.com/security/compliance/iso-27017
ISO/IEC 27018:2019 — Protection of PII	https://cloud.google.com/security/compliance/iso-27018
ISO/IEC 27701:2019 — Privacy Information Management	https://cloud.google.com/security/compliance/iso-27701
SOC 2 Type II	https://cloud.google.com/security/compliance/soc-2
CSA STAR Level 2	https://cloud.google.com/security/compliance/csa-star
PCI DSS v4.0	https://cloud.google.com/security/compliance/pci-dss