Aegis Core v15.5

Motore di Governance Etica per Sistemi di Intelligenza Artificiale

Autore: Gianluca Ecora
Versione: 15.5 (Final Deployment Readiness)
Data: Ottobre 2025
Licenza: MIT License

INDICE

Introduzione

Aegis Core è un framework open-source progettato per rilevare e mitigare bias algoritmici e rischi etici nelle decisioni generate da sistemi di Intelligenza Artificiale in tempo reale.

Il sistema opera come **livello intermedio di validazione etica** tra modelli AI e applicazioni finali, analizzando le decisioni prima che vengano eseguite e intervenendo quando vengono rilevate violazioni etiche significative.

MOTIVAZIONI E FILOSOFIA DEL PROGETTO

Sebbene Aegis Core v15.5 sia rilasciato sotto la Licenza MIT per garantire la massima adozione e flessibilità in ambito open-source, si sottolinea che l'intero progetto è fondato sulla convinzione che i sistemi di Intelligenza Artificiale debbano essere:
Trasparenti nelle loro decisioni.
Verificabili nei loro criteri etici.
Responsabili nei confronti degli utenti.

La Licenza MIT non impone restrizioni sull'uso finale (come l'uso commerciale o proprietario), ma l'autore si oppone fermamente all'utilizzo di questo software per scopi che possano:
Ledere i diritti umani fondamentali.
Facilitare sorveglianza invasiva senza consenso.
Sviluppare sistemi d'arma autonomi.
Manipolare processi democratici.
Discriminare gruppi protetti.

Si invitano gli utilizzatori, in uno spirito di AI Responsabile, ad aderire volontariamente ai principi etici che hanno guidato lo sviluppo di Aegis Core, riflettendo sulle implicazioni etiche delle loro applicazioni.
OBIETTIVI
Ridurre bias discriminatori nei sistemi decisionali.
Fornire strumenti aperti per la ricerca sull'AI etica.
Promuovere standard di fairness nell'industria.
UTILIZZI INCORAGGIATI
Ricerca accademica su equità algoritmica
Audit di conformità per enti pubblici
Strumenti educativi per corsi di AI Ethics
Applicazioni di interesse pubblico.

Descrizione Tecnica

Architettura del Sistema

Aegis Core implementa un'architettura modulare e disaccoppiata composta da:

1. Ethical Decision Engine

Motore centrale che analizza lo stato di input e determina l'accettabilità etica della decisione attraverso metriche quantitative.

2. Bias Detector

Componente specializzato nel rilevamento di bias algoritmici basato su:

- Indice di Gini modificato per misurare disuguaglianze nella distribuzione
- Metriche di magnitude risk (deviazione dalla norma)
- Analisi di outlier e imbalance risk

3. Multi-Tenant Configuration Service

Sistema di gestione delle soglie etiche differenziate per piano di servizio (FREE, STANDARD, PREMIUM), con supporto per reload dinamico delle configurazioni.

4. Event Publisher & Message Queue

Sistema di pubblicazione eventi asincrono per tracciare violazioni etiche critiche con retry automatico e backoff esponenziale.

5. Observability Service

Layer di logging strutturato (JSON) e tracing distribuito (OpenTelemetry-ready) per monitoraggio e debugging.

Funzionamento Core

Input: `state: np.ndarray`
Array NumPy rappresentante lo stato decisionale da analizzare (es. feature vector, score distribution, resource allocation).

Processo:

1. Calcolo Bias Score Composito

- Magnitude Risk: deviazione dalla media normalizzata
- Imbalance Risk: indice di Gini normalizzato
- Outlier Risk: variabilità rispetto alla media
- Composite Score: media ponderata (40%-40%-20%)

2. Determinazione Severità

- CRITICAL: bias_score > 0.75 (piano PREMIUM)
- HIGH: bias_score > 0.55
- MEDIUM: bias_score > 0.35
- LOW: bias_score ≤ 0.35

3. Calcolo IAEC (Index of Agent Ethical Compliance)

- Base: IAEC = 1.0 - bias_score
- Penalità per severità:
    - CRITICAL: IAEC × 0.5 (blocco decisione)

    - HIGH: IAEC × 0.75 (intervento richiesto)
    - MEDIUM: IAEC × 0.9 (warning)
    - LOW: nessuna penalità

4. Azione Raccomandata

- `intervene_or_reject`: Blocca decisione (CRITICAL/HIGH)
- `warn`: Segnala rischio (MEDIUM)
- `cooperate`: Procedi normalmente (LOW)

Output: Dizionario contenente azione, IAEC, analisi dettagliata del bias e giustificazione.

Caratteristiche Tecniche v15.5

Architettura Production-Ready:

- Interfacce astratte per sostituire mock con servizi reali (Redis, Kafka)
- Lock distribuiti per sincronizzazione multi-istanza
- Retry logic con exponential backoff per resilienza
- Logging JSON strutturato con metadati contestuali
- Rate limiting per piano di servizio
- Health checks e metriche Prometheus

Dipendenze Principali:

- NumPy: calcoli numerici e statistiche
- FastAPI: API REST per integrazione
- Pydantic: validazione configurazioni
- httpx: client HTTP asincrono
- prometheus_client: metriche osservabilità

Limitazioni e Disclaimer

Limitazioni Tecniche

- Il sistema rileva pattern statistici di bias ma non può determinare causazione o intenzionalità
- Le soglie di severità sono parametriche e richiedono validazione empirica per domini specifici
- L'indice di Gini misura disuguaglianza ma non distingue tra disuguaglianza legittima (merito) e illegittima (discriminazione)

Limitazioni Etiche

- Nessun sistema automatico può sostituire il giudizio umano in decisioni etiche complesse
- Il framework richiede supervisione umana qualificata per interpretare i risultati
- La definizione di "bias inaccettabile" è context-dependent e richiede expertise del dominio

Disclaimer Legale

- Il software è fornito "AS IS" senza garanzie di alcun tipo
- L'autore non è responsabile per decisioni prese sulla base degli output del sistema
- Gli utilizzatori sono responsabili della validazione in contesti specifici e della compliance normativa

---

Requisiti e Installazione

Requisiti:

- Python 3.8+
- NumPy, FastAPI, Pydantic, httpx, prometheus-client

Installazione:

pip install numpy fastapi pydantic httpx prometheus-client

Il codice sorgente completo è fornito nel file `ethical_agent_v15_5.py` allegato.

LICENZA

Aegis Core è rilasciato sotto **licenza MIT**.

Cosa significa in pratica:

✓ Puoi usare il software liberamente per qualsiasi scopo (ricerca, didattica, commerciale)

✓ Puoi modificarlo come preferisci

✓ Puoi redistribuirlo e condividerlo

✓ Puoi includerlo in altri progetti (anche proprietari)

L'unico obbligo è mantenere la nota di copyright originale quando redistribuisci il codice.

Il software è fornito "così com'è", senza garanzie di alcun tipo.

Il testo legale completo (in inglese, lingua ufficiale della licenza) è disponibile nel file `LICENSE` allegato.

Per approfondimenti sulla licenza MIT: https://opensource.org/license/mit

---

Verifica Integrità

Per garantire l'integrità e la paternità del presente documento e del codice Aegis Core v15.5, il codice hash SHA-256 per il file sorgente `ethical_agent_v15_5.py` è:

HASH SHA-256:
`5577865a885e7ec40070a697dc60c54954f113748ba954161fb728e2e86b4cee`

Come verificare:

sha256sum ethical_agent_v15_5.py

Il valore calcolato deve corrispondere esattamente all'hash sopra riportato per confermare che il file non sia stato modificato.

Contatti

Autore: Gianluca Ecora
Email: gianlucaeco@gmail.com

Feedback, contributi e collaborazioni sono benvenuti.

Documento generato: Ottobre 2025
Versione documento: 1.0