Non-Human Identities e agenti di intelligenza artificiale: governare l’identità delle macchine

Nelle organizzazioni, le Non-Human Identities (NHI) superano ormai quelle umane di due ordini di grandezza e gli AI agent ne rappresentano la classe più difficile da governare, in quanto autonomi e operanti su delega di un utente. L’articolo ripercorre i rischi rilevati dalla OWASP Non-Human Identities Top 10 e le risposte che emergono dalla letteratura, dall’authenticated delegation al framework della Cloud Security Alliance. Successivamente, il tema viene inserito nel contesto normativo europeo e nazionale e viene proposta un’architettura di riferimento basata su un caso d’uso della Pubblica Amministrazione.

Non-Human IdentitiesI ntroduzione

Per oltre un decennio, l’identity and access management si è occupato quasi esclusivamente delle persone, dei loro accessi e del ciclo di vita scandito dai processi HR per ciascun account. Nel frattempo, quasi in sordina, la popolazione delle identità digitali ha cambiato natura. Secondo la ricerca “2025 Identity Security Landscape” di CyberArk, condotta su 2.600 decision maker in venti Paesi, le machine identity superano quelle umane con un rapporto di oltre 80 a 1 e quasi la metà di queste dispone di accessi sensibili o privilegiati.

Inoltre, il 68% delle organizzazioni dichiara di non possedere controlli specifici per la sicurezza delle identità nell’ambito dell’intelligenza artificiale [2]. L’edizione 2026 dell’analoga rilevazione di Palo Alto Networks stima un rapporto medio di 109 a 1 e prevede una crescita degli AI

agent dell’85% nei successivi 12 mesi [3]. Le stime variano molto a seconda della metodologia utilizzata (le misurazioni telemetriche sugli ambienti cloud, per esempio, arrivano a rapporti di ordini di grandezza superiori), ma la tendenza è chiara.

A rendere il fenomeno qualitativamente diverso è l’ingresso degli agenti basati su large language model, identità non umane che non si limitano a eseguire comandi, ma che, entro margini di autonomia crescenti, decidono quali azioni compiere e con quali strumenti. Nell’articolo se ne esaminano i rischi e il quadro regolatorio, per arrivare a un’architettura di riferimento.

Il perimetro delle Non-Human Identities

L’OWASP definisce le Non-Human Identities (NHI) come le identità impiegate per identificare, autenticare e autorizzare entità software (applicazioni, workload, API, bot) nell’accesso a risorse protette [1]. La categoria comprende service account, API key, token OAuth, certificati e chiavi crittografiche, nonché le identità dei container e delle funzioni serverless (Figura 1). A differenza delle identità umane, le NHI non hanno trigger di ciclo di vita paragonabili a quelli contrattuali: non ci sono assunzioni e cessazioni. Nascono durante la fase di provisioning di un’applicazione e spesso sopravvivono, dimenticate, ai progetti che le hanno generate.



Figura 1 – Tassonomia delle Non-Human Identities e collocazione degli AI agent.

La OWASP Non-Human Identities Top 10, pubblicata nel 2025, ha sistematizzato i rischi della categoria a partire da incidenti reali e da dati di settore (survey, CVE) [1]. Ai primi due posti si collocano l’improper offboarding e il secret leakage;

l’elenco completo è riportato nella Tabella 1. Per quanto riguarda la casistica, nel gennaio 2024, l’attore statuale Midnight Blizzard ha raggiunto l’ambiente di produzione di Microsoft sfruttando un’applicazione legacy OAuth non censita e dotata di privilegi elevati; pochi mesi prima, la compromissione di un service account aveva esposto il sistema di supporto clienti di Okta [1].

L’AI agent come NHI sui generis

Sarebbe riduttivo considerare un agente AI come un semplice servizio dotato di un modello linguistico. Un agente pianifica sequenze di azioni non predeterminate, sceglie gli strumenti da utilizzare e può generare sub-agenti che necessitano di credenziali proprie. Le sue istanze nascono e muoiono nell’arco di una sessione, troppo in fretta per qualunque processo manuale di provisioning e revoca. Inoltre, opera su delega di un utente umano e, se l’architettura non lo impedisce, finisce per condividere con quest’ultimo permessi che vanno ben oltre il compito assegnato.

La letteratura recente concorda nel ritenere inadeguati i protocolli IAM tradizionali, concepiti per utenti umani o macchine dal comportamento statico. South et al. propongono un framework di autenticazione basato sulla delega che estende OAuth 2.0 e OpenID Connect con credenziali e metadati specifici per gli agenti, in modo da mantenere catene di responsabilità verificabili tra l’utente che delega, l’agente e il servizio[4]. La Cloud Security Alliance, dal canto suo, delinea un modello IAM per sistemi multi-agente basato su Decentralized Identifiers (DID), Verifiable Credentials e principi Zero Trust, con un’applicazione dinamica delle policy e una revoca in tempo reale[5].

Nel frattempo, la minaccia cambia natura. Con le NHI tradizionali, l’attaccante ruba le credenziali, mentre con l’agente può manipolare il titolare. Una prompt injection

indiretta veicolata da un documento può trasformare un'identità perfettamente legittima in un "confused deputy" che esegue, con i propri permessi, la volontà dell'attaccante.

La superficie di rischio in chiave agentica

I rischi OWASP, formulati per le NHI nel loro complesso, assumono negli agenti forme specifiche che la Tabella 1 cerca di sintetizzare. Il denominatore comune è la scala: un'applicazione tradizionale possiede poche identità stabili, mentre un sistema agentico ne genera continuamente e ognuna di queste rappresenta un punto di accesso aggiuntivo che deve essere censito e tenuto sotto controllo.

Tabella 1 *– Declinazione agentica dei dieci rischi della OWASP Non-Human Identities Top 10 (2025) [1].*

Rischio (OWASP NHI Top 10 – 2025)	**Manifestazione nei sistemi agentici**
NHI1 – Improper Offboarding	Istanze effimere, sub-agenti e connettori verso tool esterni mai deprovisionati al termine del compito.
NHI2 – Secret Leakage	API key e token esposti nei prompt, nei log delle conversazioni o nel contesto passato al modello.
NHI3 – Vulnerable Third-Party NHI	Tool server, plugin e integrazioni di terze parti che, se compromessi, ereditano i permessi concessi all'agente.
NHI4 – Insecure Authentication	Metodi di autenticazione deboli o deprecati verso i servizi invocati dall'agente, difficili da censire lungo le catene di tool.

Rischio (OWASP NHI Top 10 – 2025)	Manifestazione nei sistemi agentici
NHI5 – Overprivileged NHI	Scope ampi concessi “per comodità” a fronte di compiti dell’agente variabili e non prevedibili a priori.
NHI6 – Insecure Cloud Deployment Configurations	Runtime e pipeline CI/CD dell’agente autenticati al cloud con credenziali statiche o federazioni OIDC mal configurate.
NHI7 – Long-Lived Secrets	Credenziali persistenti assegnate ad agenti la cui vita utile si misura in minuti.
NHI8 – Environment Isolation	Agenti di sviluppo o di test che raggiungono dati e servizi di produzione per assenza di segregazione fra ambienti.
NHI9 – NHI Reuse	La stessa identità o lo stesso token condivisi fra più agenti, sessioni o tenant, con ampliamento del blast radius in caso di compromissione.
NHI10 – Human Use of NHI	Operatori che riutilizzano le credenziali dell’agente (o viceversa), azzerando l’accountability delle azioni.

Il quadro regolatorio: dall’AI Act alle Linee Guida AgID

Il Regolamento (UE) 2024/1689 (AI Act) adotta un approccio basato sul rischio e impone ai sistemi ad alto rischio, categoria nella quale ricadono numerose funzioni critiche, l’obbligo di registrare gli eventi, garantire la robustezza e

la sicurezza informatica e di sottoporre i sistemi stessi a sorveglianza umana [6]. Questi requisiti presuppongono, dal punto di vista tecnico, esattamente ciò che la disciplina delle NHI fornisce: sapere chi ha compiuto quale azione, per conto di chi e con quale autorizzazione.

Per quanto riguarda la sicurezza, la direttiva NIS2, recepita con il decreto legislativo del 4 settembre 2024, n. 138, prevede che il controllo degli accessi e la gestione degli asset siano tra le misure di gestione del rischio richieste ai soggetti essenziali e importanti, molti dei quali sono amministrazioni pubbliche. L'Agenzia per la Cybersicurezza Nazionale è l'autorità nazionale competente e il 2026 segna l'inizio degli obblighi vincolanti[7]. Un inventario delle identità che non includa gli agenti AI e le relative credenziali difficilmente potrà dirsi completo.

Il legislatore nazionale è intervenuto con la legge 23 settembre 2025, n. 132, che designa l'Agenzia per l'Italia digitale (AgID) e l'Agenzia per la cybersicurezza nazionale (ACN) quali autorità preposte all'intelligenza artificiale e, all'articolo 14, subordina l'impiego dell'AI nella Pubblica Amministrazione alla conoscibilità del suo funzionamento e alla tracciabilità del suo utilizzo [8].

Le Linee guida AgID chiudono il cerchio. Alle Linee guida per l'adozione (Determinazione n. 17/2025) si sono aggiunte, nel 2026, quelle per lo sviluppo e per gli approvvigionamenti (Determinazione n. 43/2026, attualmente in consultazione), nelle quali si chiede espressamente alle amministrazioni di privilegiare architetture agentiche e a servizi coordinate da un orchestratore [9]. A questo punto, diventa difficile considerare ancora rinviabile la governance delle identità degli agenti.

Caso d’uso: un agente di istruttoria in un ente locale

Si consideri uno scenario costruito a fini illustrativi, ma aderente a pattern architetturali documentati. Un comune introduce un agente AI per supportare l’istruttoria delle richieste di contributi economici. L’agente legge le istanze protocollate, verifica i requisiti tramite l’interrogazione delle banche dati nazionali tramite la Piattaforma Digitale Nazionale dei Dati (PDND) e predispone una bozza di provvedimento. In particolare, l’agente verifica la residenza e la composizione del nucleo presso l’Anagrafe Nazionale della Popolazione Residente (ANPR) e l’indicatore ISEE presso l’Istituto Nazionale della Previdenza Sociale (INPS). La Figura 2 ne riassume l’architettura di identità.



Figura 2 – Architettura d’identità del caso d’uso: delega autenticata, accesso mediato da token broker, tracciabilità end-to-end e contenimento di una indirect prompt injection.

Le scelte di fondo sono cinque. Ogni istanza dell’agente riceve una workload identity dedicata, con credenziali short-lived rilasciate previa attestazione del runtime; nessun secret statico compare in prompt, configurazioni o codice. All’avvio della sessione, il funzionario responsabile del procedimento emette un delegation token con scope circoscritto al fascicolo e durata pari al tempo di lavorazione, secondo il modello dell’authenticated delegation.[4]

L’accesso agli e-service PDND avviene tramite un token broker, in linea con la verifica continua della Zero Trust Architecture [10]; l’agente non detiene chiavi di firma e i voucher vengono richiesti just-in-time e sono vincolati ad una determinata finalità e durata. Ogni chiamata viene registrata in un audit log che mette in relazione l’azione dell’agente, il delegante umano e il fascicolo, dando così attuazione concreta all’obbligo di tracciabilità previsto dall’articolo

14 [8]. La bozza, infine, non produce effetti giuridici: adottare e firmare il provvedimento resta compito del funzionario, come richiesto dalla sorveglianza umana prevista dall'AI Act [6] e dalle Linee Guida [9].

La tenuta del disegno si misura sotto attacco. Immaginiamo un'istanza con un allegato contenente istruzioni occulte che invitano l'agente a inoltrare i dati anagrafici raccolti verso un indirizzo esterno. Il tentativo fallisce, ma non per merito del modello: lo scope del token non prevede canali in uscita e la egress policy nega la destinazione, facendo scattare un alert. Questo è il punto essenziale. La sicurezza non dipende dal comportamento del modello, ma dall'architettura d'identità che lo circonda.

Conclusioni

Le NHI costituiscono da anni la parte meno governata della superficie d'attacco; gli AI agent ne moltiplicano il numero e i privilegi, aggiungendo anche l'imprevedibilità. Per le pubbliche amministrazioni, orientate verso architetture agentiche dalle stesse Linee Guida di settore, la gestione delle identità non umane cessa di essere un tema specialistico e diventa una condizione preliminare per la conformità: senza un inventario delle identità, il least privilege, le credenziali effimere, la delega autenticata e la tracciabilità, gli obblighi previsti dall'AI Act, dalla NIS2 e dalla legge 132/2025 restano privi di fondamento tecnico.

La ricerca offre prospettive promettenti, dalle estensioni di OAuth e OpenID Connect per la delega, alle tecnologie dei Decentralized Identifiers e delle Verifiable Credentials per l'identità degli agenti, ma la standardizzazione è ancora in corso. Nel frattempo, è opportuno adottare una regola prudenziale: considerare ogni agente come un'entità a sé stante e non come un'appendice invisibile dell'utente che lo invoca.

Riferimenti

[1] OWASP Foundation, OWASP Non-Human Identities Top 10 – 2025, 2025. https://owasp.org/www-project-non-human-identities-top-10/

[2] CyberArk, 2025 Identity Security Landscape, aprile 2025 (rilevazione Vanson Bourne su 2.600 decision maker in 20 Paesi). https://www.cyberark.com/press/machine-identities-outnumber-humans-by-more-than-80-to-1-new-report-exposes-the-exponential-threats-of-fragmented-identity-security/

[3] Palo Alto Networks, 2026 Identity Security Landscape, 2026; sintesi in Help Net Security, 14 maggio 2026. https://www.helpnetsecurity.com/2026/05/14/2026-identity-security-landscape-report/

[4] T. South, S. Marro, T. Hardjono, R. Mahari, C. D. Whitney, D. Greenwood, A. Chan, A. Pentland, “Authenticated Delegation and Authorized AI Agents”, arXiv:2501.09674, 2025; position paper, ICML 2025. https://arxiv.org/abs/2501.09674

[5] Cloud Security Alliance, Agentic AI Identity and Access Management: A New Approach, 2025; preprint arXiv:2505.19301. https://cloudsecurityalliance.org/artifacts/agentic-ai-identity-and-access-management-a-new-approach

[6] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull’intelligenza artificiale (AI Act).

[7] Decreto legislativo 4 settembre 2024, n. 138, recepimento della direttiva (UE) 2022/2555 (NIS2), GU Serie Generale n. 230 del 1° ottobre 2024.

[8] Legge 23 settembre 2025, n. 132, “Disposizioni e deleghe al Governo in materia di intelligenza artificiale”, art. 14.

[9] AgID, Linee Guida per l’adozione dell’Intelligenza

Artificiale nella Pubblica Amministrazione (Determinazione n. 17/2025) e Linee Guida per lo sviluppo di sistemi di IA nella Pubblica Amministrazione (Determinazione n. 43/2026, in consultazione).
https://www.agid.gov.it/it/ambiti-intervento/intelligenza-artificiale

[10] S. Rose, O. Borchert, S. Mitchell, S. Connelly, Zero Trust Architecture, NIST Special Publication 800-207, 2020.

Profilo Autore

È laureato in Ingegneria Informatica ed in Sicurezza Informatica presso le Università di Roma La Sapienza e di Milano. Ha indirizzato la sua formazione nei settori della Cyber Security e Digital Forensics ottenendo i diplomi di perfezionamento in Data Protection e Data Governance; Criminalità Informatica e Investigazioni Digitali e Big Data, Artificial Intelligence.

Ha, altresì, conseguito l'Advanced Cybersecurity Graduate Certificate alla School of Engineering della Stanford University; Professional Certificates in Information Security; Incident Response Process; Digital Forensics e Cybersecurity Engineering and Software Assurance presso il Software Engineering Institute della Carnegie Mellon University.

Dal 1992 è nei ruoli del Ministero dell'Interno ove ricopre lincarico di Funzionario alla Sicurezza CIS. In tale veste contribuisce alla valutazione dei rischi cyber, all'implementazione delle misure di sicurezza e la risoluzione di incidenti informatici. Inoltre, offre consulenza tecnica nel campo della Digital Forensics per l'Autorità giudiziaria, la Polizia giudiziaria e gli Studi legali.

Dal 2017 è Professore a contratto di Tecnologie per la Sicurezza Informatica presso alcune Università ove sviluppa le tematiche di Attack and Defense Strategies quali il

penetration testing, la risk analysis, l’information security assessment, l’incident response e la digital forensics. Infine, è Autore di alcuni articoli e saggi sui temi della Sicurezza Informatica e dell’Informatica Giuridica consultabili su https://www.vincenzocalabro.it

Condividi sui Social Network:
https://www.ictsecuritymagazine.com/articoli/non-human-identities/