Digeat N.6 - 19 Giugno 2025

L’autenticità riconfigurata. Governance digitale tra cybersecurity rafforzata, integrità informativa e rischi inediti dell’IA

Di Enrica Priolo

Abstract

L’infosfera digitale contemporanea è segnata da una crisi epistemologica legata all’erosione dell’autenticità. Questo articolo evidenzia come la nozione di autenticità trascenda la sua tradizionale accezione archivistica per diventare una qualità del sistema trasversale, la cui garanzia dipende da un equilibrio complesso tra innovazione digitale, cybersecurity rafforzata e trasparenza dei procedimenti di creazione delle informazioni. Analizzando il contributo del Cybersecurity Act e l’impatto della Direttiva NIS2, nonché considerando i rischi inediti posti dall’IA generativa alla veridicità dell’informazione, si propone un framework operativo – il ciclo V.I.T.A.– come modello di governance per imprese e professionisti, volto a costruire e mantenere un’autenticità di processo difendibile nell’ecosistema digitale.

Indice

La crisi epistemologica dell’infosfera digitale

Ci troviamo immersi in quella che potremmo definire, parafrasando Lyotard, una condizione post-moderna accelerata dalla tecnologia, caratterizzata da una frammentazione delle grandi narrazioni e da una crescente difficoltà nel discernere il segnale dal rumore, l’informazione attendibile dalla manipolazione deliberata o dall’errore involontario. **Il dominio dell’autenticità**, inteso come quello spazio intersoggettivo e fiduciario basato su fonti riconoscibili e processi verificabili, che rende possibile la comunicazione significativa, la conoscenza condivisa e la deliberazione democratica, **è oggi sottoposto a una pressione senza precedenti.**

> La digitalizzazione pervasiva, pur avendo democratizzato l’accesso all’informazione, ha simultaneamente creato un ambiente in cui la velocità virale, l’opacità algoritmica dei sistemi di raccomandazione (spesso ottimizzati per l’engagement a scapito dell’accuratezza) e, più recentemente, l’irruzione dell’Intelligenza Artificiale generativa

> con la sua capacità mimetica quasi perfetta, minano le nostre tradizionali àncore epistemiche.

Non si tratta solo della proliferazione di fake news, ma di **una più profonda crisi di fiducia nelle fonti, nelle istituzioni e negli stessi artefatti informativi che mediano la nostra realtà.**

In questo scenario complesso, il legislatore europeo ha reagito non con interventi estemporanei, ma cercando di **costruire un ecosistema normativo multilivell**o, volto a rafforzare la resilienza e l'affidabilità dell'infrastruttura digitale. Il Regolamento UE 2019/881 -che istituisce un quadro per la certificazione della cybersicurezza- e la Direttiva UE 2022/2557 -che innalza drasticamente gli obblighi di sicurezza e di governance per un'ampia platea di soggetti essenziali e importanti, rappresentano pilastri fondamentali di questa strategia.

Osservo come tali normative, unitamente ai principi della protezione dati e alle prime regole sull'AI Act, contribuiscano ad una riconfigurazione necessaria del concetto di autenticità. **Sostengo che l'autenticità, nell'era digitale, non può più essere vista come una proprietà statica, bensì come una qualità "processuale" emergente**, il risultato di una governance integrata in cui la cybersecurity gioca un ruolo infrastrutturale abilitante. Proporrò, infine, un framework operativo – il ciclo V.I.T.A. – come modello per le organizzazioni che intendono perseguire e dimostrare questa nuova forma di autenticità processuale.

La metamorfosi del concetto di autenticità

La concezione classica dell'autenticità, radicata nella tradizione archivistica, diplomatica e filologica, si fondava sulla verificabilità dell'origine (*provenance*), sull'integrità materiale del supporto e sulla conformità a sigilli, firme o stili riconosciuti. La digitalizzazione polverizza questi ancoraggi.

> L'informazione diventa fluida, separabile dal suo supporto originario, infinitamente riproducibile (spesso senza distinzione tra originale e copia), facilmente alterabile e ricontestualizzabile. L'IA generativa aggiunge un ulteriore livello di complessità, potendo creare *ex novo* contenuti che imitano perfettamente stili e formati cd. autentici, ma privi di un referente reale o di un'intenzionalità autoriale umana trasparente.

Questo impone un ripensamento ontologico ed epistemologico. **Se l'autenticità non può più risiedere unicamente nell'oggetto informativo in sé, essa deve essere ricercata nel processo che lo genera, lo gestisce, lo trasmette e lo rende accessibile.** L'autenticità diventa una qualità relazionale e contestuale, legata all'affidabilità della catena informativa. Come suggerito dai filosofi della tecnologia, l'etica dell'informazione si sposta verso la valutazione della qualità dei processi e degli ambienti informativi. In questa prospettiva, potremmo definire l'autenticità processuale come la risultante verificabile di quattro dimensioni interconnesse.

Innanzitutto deve esistere la capacità di risalire in modo affidabile all'origine (umana o algoritmica) dell'informazione; inoltre, è necessaria la ragionevole certezza che l'informazione non sia stata alterata in modo non autorizzato durante il suo ciclo di vita; ancora, serve la chiarezza riguardo alle finalità, ai limiti e alle condizioni di produzione e diffusione dell'informazione ed, infine, l'esistenza di un soggetto (persona fisica o giuridica) identificabile e responsabile per la correttezza e la legittimità dell'informazione.

> Questa visione di processo non nega l'importanza della corrispondenza fattuale (verità), ma riconosce che, nell'infosfera complessa, la fiducia (*trust*) si costruisce

> primariamente sulla base dell'affidabilità, percepita e verificabile, dei processi e delle fonti, anziché sulla verifica diretta di ogni singolo enunciato.

Il ruolo fondante della cybersecurity rafforzata

In questo quadro, **la cybersecurity cessa di essere una mera disciplina tecnica di difesa perimetrale per assurgere a funzione infrastrutturale abilitante dell'autenticità processuale.** La sua capacità di garantire l'integrità, la riservatezza (spesso precondizione per l'autenticità della comunicazione) e la disponibilità dei sistemi informativi è la base tecnica su cui poggia la fiducia nell'ecosistema digitale. Il Cybersecurity Act contribuisce direttamente attraverso il quadro europeo di certificazione della cybersecurity. Gli schemi di certificazione (Art. 54 CSA), basati su livelli di affidabilità differenziati (base, sostanziale, elevato), forniscono un marchio di affidabilità verificabile per prodotti, servizi e processi ICT. Utilizzare componenti certificati CSA, in un flusso informativo, ne aumenta la presunzione di integrità e sicurezza, fornendo un segnale di affidabilità sulla *provenienza* (se la certificazione riguarda l'identità del produttore) e sull'*integrità* del sistema che tratta l'informazione.

Dal canto suo, invece, **la Direttiva NIS2 (e la sua attuazione nazionale) opera in modo ancora più pervasivo**, imponendo obblighi di governance e gestione del rischio cyber a un'ampia platea di organizzazioni. Le misure minime di sicurezza richieste dall'art. 21 della NIS2 sono direttamente funzionali a garantire l'autenticità processuale. Basti pensare alle policy di analisi dei rischi e dei sistemi informativi, alla gestione degli incidenti, alla *business continuity*, alla gestione della crisi, alla sicurezza della catena di approvvigionamento, alla sicurezza nell'acquisizione, sviluppo e manutenzione di sistemi, alla promozione della *security by design* in genere, alle pratiche di igiene informatica e formazione, fino alla crittografia, alla *encryption* e all' MFA.

> L'obbligo per gli organi di gestione di approvare e supervisionare queste misure (previsto dalla normativa di recepimento come per l'Art. 32 Direttiva UE 2016/1148 e rafforzato da NIS2) rende la cybersecurity un tema di alta direzione, intrinsecamente legato alla responsabilità complessiva per l'affidabilità e l'autenticità dei processi aziendali.

Sulla IA come fabbrica potenziale di inautenticità

L'IA generativa rappresenta, oggi, il vettore più potente di sfida all'autenticità processuale. Modelli come GPT-4 o DALL-E 3 generano contenuti la cui origine è puramente algoritmica, ma che possono essere presentati come opera umana o attribuiti falsamente. La capacità di imitare stili specifici rende ardua l'identificazione della fonte reale. Oltre alla creazione *ex novo* di *deepfake* (audio, video, immagini), l'IA può essere usata per alterazioni sottili e mirate di contenuti esistenti (es. modificare leggermente un virgolettato in un articolo, alterare un dettaglio in una foto), difficilmente rilevabili ad occhio nudo o con strumenti tradizionali.

> L'IA può essere usata per generare contenuti apparentemente neutrali o informativi, ma progettati per manipolare l'opinione, diffondere narrazioni di parte o promuovere interessi nascosti su scala industriale e con personalizzazione estrema.

I modelli generativi possono produrre informazioni fattualmente errate, ma linguisticamente fluenti (le "allucinazioni"), che, se non verificate, entrano nel circolo informativo come dati apparentemente autentici, inquinando la base di conoscenza condivisa.

L’AI Act tenta di rispondere con gli obblighi di trasparenza dell’Art. 52: chi usa sistemi IA per generare o manipolare contenuti immagine, audio o video che assomiglino notevolmente a persone, oggetti, luoghi esistenti in modo ingannevole (*deep fake*), deve dichiararne la natura artificiale o manipolata. Similmente, chi interagisce con sistemi IA (come le chatbot) deve essere informato di tale interazione. Questi obblighi sono un passo necessario, ma la loro efficacia pratica dipende dalla possibilità tecnica di implementare meccanismi di *watermarking* robusti e resistenti alla manomissione, nonché dalla capacità degli utenti di riconoscere e interpretare correttamente tali segnalazioni. La sfida della *detection* e del *labeling* affidabile su scala globale rimane aperta.

Ciclo V.I.T.A., un possibile framework di partenza per difendere l’autenticità di processo in azienda

Per affrontare questa complessità in modo strutturato, **propongo un framework di governance aziendale denominato Ciclo V.I.T.A. (Verifica, Integrità, Tracciabilità, Accountability).** Nulla di originale, ma semplicemente un modo di ragionare in termini di gestione del rischio, così come ci ha abituato il legislatore europeo.

Non un insieme rigido di controlli, ma **un approccio ciclico basato sui principi per costruire e mantenere l’autenticità di processo delle informazioni critiche**, partendo dalla validazione delle fonti e degli input, passando per la protezione dell’informazione e per la documentazione della provenienza, per arrivare ad una responsabilità chiara e a rimedi efficaci.

Verifica:

- *Identità della fonte*. Implementare meccanismi robusti per verificare l’identità di chi crea o fornisce informazioni (umani o sistemi), utilizzando autenticazione forte (MFA), identità digitali federate o certificate (eIDAS) o firme digitali qualificate.
- *Affidabilità dei componenti.* Privilegiare l’uso di prodotti e servizi ICT con certificazioni di sicurezza (CSA) o valutazioni indipendenti, specialmente per le infrastrutture critiche.
- *Validazione dei dati in ingresso.* Stabilire controlli (automatici e/o manuali) sulla plausibilità, coerenza e qualità dei dati utilizzati nei processi informativi, specialmente quelli usati per addestrare o alimentare sistemi IA. *Due diligence* rafforzata su *dataset* di terze parti.
- *Approccio Zero Trust (esteso all’Informazione).* Non fidarsi implicitamente di nessuna fonte, ma verificare continuamente l’identità e l’autorizzazione all’accesso/modifica dell’informazione.

Integrità:

- *Misure di sicurezza NIS2-aligned.* Implementare l’intero spettro delle misure tecniche e organizzative richieste dall’art. 21 della NIS2 per proteggere l’informazione da alterazioni non autorizzate: sicurezza delle reti, gestione delle vulnerabilità, crittografia (*end-to-end*, *at rest*), controlli di accesso rigorosi (RBAC, ABAC, PAM), *Data Loss Prevention* (DLP).
- *Garanzie crittografiche.* Utilizzo sistematico di funzioni di *hash* per verificare l’integrità dei file, sigilli elettronici qualificati (eIDAS) per garantire integrità e origine, *timestamping* qualificato per la data certa.
- *Sicurezza nello sviluppo (DevSecOps).* Integrare controlli di sicurezza e integrità fin dalla fase di sviluppo del *software* e dei modelli IA (es. tecniche di *robust training* contro *adversarial attacks*).
- *Gestione controllata delle modifiche.* Procedure rigorose di *change management* per qualsiasi modifica a dati o sistemi critici, con tracciatura e approvazione.

Tracciabilità:

- *Audit trail immutabili e dettagliati.* Implementare sistemi di *logging* sicuri (WORM-like), che registrino in modo non ripudiabile chi ha creato, acceduto, modificato o cancellato informazioni critiche, quando e come.
- *Metadati sulla provenienza.* Adottare (o contribuire a definire) standard di metadati (es. C2PA) che accompagnino i contenuti digitali indicandone l'origine (umana o IA), le modifiche subite e gli strumenti utilizzati.
- *Trasparenza sulla generazione AI (AI Act Art. 52).* Implementare meccanismi tecnici (es. *watermarking* robusto) e procedurali per segnalare in modo chiaro e persistente i contenuti generati o significativamente alterati dall'IA.
- *Data Lineage.* Ove possibile, mappare il flusso dei dati attraverso i sistemi per comprenderne l'origine e le trasformazioni subite, essenziale per *l'accountability* e il *debugging* (specie per l'IA).

Accountability:

- *Attribuzione di responsabilità.* Definire chiaramente nei processi e nelle *policy* interne chi è responsabile per la verifica, l'integrità, la tracciabilità e l'accuratezza delle informazioni in ogni fase. La responsabilità ultima risale agli organi di gestione (cfr. NIS2).
- *Policy interne dettagliate.* Formalizzare le procedure V.I.T.A. in *policy* accessibili e comunicate a tutto il personale.
- *Meccanismi di segnalazione e correzione.* Canali accessibili (interni/esterni) per segnalare errori, sospette manipolazioni o contenuti inautentici. Procedure definite per l'investigazione e la rettifica tempestiva.
- *Risposta agli incidenti di autenticità.* Includere scenari di crisi legati a disinformazione massiva o *deepfake* dannosi nei piani di *incident response* e *business continuity*.
- *Integrazione con compliance (GDPR, 231/2001).* Collegare le debolezze nei processi V.I.T.A. ai rischi di non conformità GDPR (es. violazione integrità Artt. 5,32) o potenzialmente alla responsabilità penale (se l'inadeguata gestione dell'autenticità agevola reati presupposto).

> Implementare il Ciclo V.I.T.A. significa trasformare l'autenticità da concetto astratto a processo gestito, misurabile e difendibile.

Implicazioni operative per imprese e professionisti

L'adozione di un approccio basato sull'autenticità di processo richiede un impegno trasversale nell'azienda; tutte le funzioni devono essere coinvolte. Ad esempio, la funzione legal & compliance dovrebbe guidare la definizione delle policy V.I.T.A., assicurare l'allineamento con GDPR, AIA, NIS2, CSA e altre normative, gestire i rischi legali associati, definire clausole contrattuali adeguate con fornitori ICT/AI. Quella IT & Cybersecurity dovrebbe implementare e gestire l'infrastruttura tecnica per la Verifica, l'Integrità e la Tracciabilità (sistemi di identità, crittografia, *logging*, sicurezza perimetrale e applicativa), collaborando strettamente con il legal & compliance. Comunicazione, marketing, R&D ed HR dovrebbero integrare i principi V.I.T.A. nei loro processi quotidiani di creazione e diffusione delle informazioni, essere consapevoli dei rischi (es. uso di IA generativa per contenuti esterni) e seguire le procedure definite.

> Inoltre, serve una formazione diffusa non solo sulla cybersecurity tecnica, ma sulla consapevolezza critica riguardo all'autenticità dell'informazione, ai rischi della disinformazione e alle procedure V.I.T.A. da applicare.

Chiaramente, un protocollo siffatto richiede investimenti mirati in tecnologie di sicurezza, strumenti di verifica, piattaforme di *logging* e, soprattutto, in competenze umane. Va però considerato quale investimento nella riduzione del rischio (legale, reputazionale, operativo) e nella costruzione di fiducia, con ritorni potenzialmente superiori ai costi.

Preservare il dominio dell'autenticità significa, in ultima analisi, preservare le fondamenta epistemiche della nostra capacità di comprendere il mondo, di comunicare significativamente e di deliberare democraticamente come società libera e informata.