POLLICINO & PARTNERS
AIDVISORY

La sicurezza cognitiva nei processi di *governance* aziendale[1]

Valerio De Luca
Fondatore e Direttore, SPES Academy "Carlo Azeglio Ciampi". Presidente, Fondazione AISES e ConnectED Mind.

Giovanni Giamminola
Founder & CEO, Systemic Zero. Senior Advisor on AI-Driven Cognitive Governance, Pollicino AIdvisory. Fellow, UCL Institute for Sustainable Business.

Oreste Pollicino
Managing Partner, Pollicino AIdvisory. Professore Ordinario di Diritto Costituzionale.

Abstract
Il contributo analizza l'intersezione tra sicurezza cognitiva e *governance* aziendale dell'intelligenza artificiale, argomentando che i sistemi di AI generativa e agentica non si limitano a introdurre rischi tecnici o reputazionali, ma esercitano un potere cognitivo che interviene sulle architetture decisionali delle organizzazioni prima ancora che i processi formali di controllo possano attivarsi. A partire da una prospettiva interdisciplinare che integra scienze cognitive, filosofia dell'informazione, diritto, teoria organizzativa e studi sulla sicurezza, il saggio propone un *framework* integrato che estende le logiche della sicurezza cognitiva al dominio della *governance* d'impresa.

[1] Il presente contributo è stato elaborato dai tre autori congiuntamente. È da attribuire specificatamente a De Luca la dimensione applicativa (i.e., sicurezza cognitiva, posizionamento istituzionale, integrazione nel contesto imprenditoriale italiano), a Giamminola il *framework* metodologico (e.g., *Thinking Quadrants, Cognitive Governance*), *field evidence,* e, insieme a Pollicino, invece, la dimensione normativa e giuridica. In particolare, è attribuibile a Pollicino l'intuizione di CogIA come strumento giuridico, *cognitive lock-out*, ed i profili di *governance* regolatoria integrata.

I. Introduzione: il problema della *governance* cognitiva nell'era dell'IA generativa

La letteratura sull'AI *governance* aziendale si è sviluppata prevalentemente attorno a categorie di rischio tecnico, compliance normativa e responsabilità giuridica. Quadri regolatori di riferimento come l'AI Act europeo, le linee guida dell'OCSE sull'IA affidabile e i principi di *corporate AI ethics* elaborati dalle principali organizzazioni internazionali condividono un'impostazione comune: identificare, classificare e mitigare i rischi associati all'*output* dei sistemi di intelligenza artificiale. Questa impostazione, pur necessaria, sconta una limitazione strutturale crescente. Essa interviene sugli effetti visibili dell'IA senza affrontare il livello in cui l'IA esercita il proprio impatto più profondo, ossia il livello pre-decisionale, dove i sistemi algoritmici strutturano il contesto cognitivo entro cui le decisioni vengono formate, prima ancora che vengano prese.

La convergenza di molteplici sviluppi recenti, provenienti da domini disciplinari normdifferenti e finora insufficientemente integrati, rende questa limitazione sempre più critica. Sul piano delle scienze cognitive, l'emergenza dell'IA generativa e dei sistemi agentici ha trasformato l'IA da strumento di supporto analitico in un'infrastruttura cognitiva pervasiva. Nella prospettiva di mappare i diversi approcci all'argomento, giova sin d'ora richiamare i diversi contributi nella letteratura scientifica che hanno trattato l'argomento e che possono fornire delle utili indicazioni di metodo per lo sviluppo di una nuova ricognizione metodologica votata all'interdisciplinarità.

In particolare, Riva (2025) ha elaborato una teoria sistematica dell'AI come infrastruttura invisibile del pensiero, mostrando come i sistemi algoritmici operino al livello delle architetture cognitive organizzative prima ancora che i decisori ne abbiano consapevolezza. Un'altra autorevole dottrina (cfr. Chiriatti, Ganapini, Panai, Ubiali e Riva, 2024) ha denominato questo fenomeno 'System 0 thinking', trattando di un livello pre-decisionale di elaborazione algoritmica che precede tanto il pensiero intuitivo (il c.d. Sistema 1 di Kahneman) quanto quello analitico (ossia il Sistema 2), strutturando il contesto cognitivo di entrambi.

Sul piano filosofico, Floridi (2010; 2014; 2019) ha elaborato un *framework* concettuale per comprendere come i sistemi digitali trasformino il substrato cognitivo delle organizzazioni, spostando irreversibilmente il valore dalla sintassi dei dati alla semantica del significato.

Inoltre, sul piano del diritto e della regolazione, si richiama il contributo di Giamminola e Pollicino (2025), che hanno prodotto il contributo teorico più avanzato disponibile nella letteratura europea, identificando la lacuna strutturale tra governance regolatoria dell'IA e *governance* cognitiva.

Infine, sul piano metodologico e operativo, Giamminola e Taticchi (2026) hanno condotto un'indagine empirica su dirigenti europei nell'arco di 18 mesi, identificando quattro posture cognitive ricorrenti nell'interazione tra leader e sistemi di IA (il c.d. Thinking Quadrants), documentando il fenomeno del *posture drift* sotto pressione decisionale e rilevando che il 43% delle organizzazioni non applica alcun meccanismo di *governance* cognitiva a livello di team. Questa analisi

empirica fornisce la base del *framework* di Cognitive AI Governance sviluppato nelle sezioni successive del presente contributo. In questo quadro sono stati elaborati sia il 'Cognitive Impact Assessment' come strumento giuridico complementare all'AI Act, sia il *cognitive lock-out* come nuova categoria giuridica che richiede un sistema di tutele specifiche nel diritto pubblico e privato.
Come anticipato, il presente saggio, alla luce della ricognizione effettuata, intende operare una sintesi interdisciplinare di queste elaborazioni, trasferendole al livello della *governance* aziendale e producendo un *framework* che consenta alle organizzazioni di governare non solo i rischi tecnici dei sistemi di IA, ma anche il potere cognitivo che essi esercitano sulle architetture decisionali interne.

II. Un *framework* di Cognitive AI Governance aziendale

Svolte le dovute premesse, prima dell'analisi di merito, occorre svolgere alcune premesse di metodo. Il *framework* di Cognitive AI Governance aziendale qui proposto integra tre filoni di ricerca convergenti: (a) il framework dei Thinking Quadrants e i relativi meccanismi di *governance* cognitiva[2] sviluppati nell'ambito della ricerca 'Systemic Zero', sopra richiamata, sulla base di 18 mesi di *fieldwork* con dirigenti europei (Giamminola, 2026; Giamminola e Taticchi, 2026); (b) la proposta normativa del 'Cognitive Impact Assessment' e la categoria giuridica del *cognitive lock-out* (Giamminola e Pollicino, 2025); infine, (c) la dimensione della sicurezza cognitiva come estensione dei *framework* di *cognitive warfare* al contesto aziendale.[3]

1. Cognitive Impact Assessment aziendale

Il primo componente operativo del framework è l'introduzione di un processo sistematico di Cognitive Impact Assessment (di seguito, CogIA), come parte integrante dei processi di adozione e gestione dei sistemi di IA in azienda. A differenza dei tradizionali *AI risk assessment* orientati agli *output*, il CogIA aziendale valuta l'impatto dei sistemi di IA sui processi cognitivi organizzativi lungo tre dimensioni: anzitutto, la dipendenza cognitiva, ossia comprendere in quale misura l'adozione del sistema riduce la capacità dell'organizzazione di elaborare autonomamente determinate categorie di problemi; in secondo luogo, il c.d. *framing effect* all'interno del quale si cerca di comprendere in quale misura il sistema AI tende a convergere verso specifici *frame* cognitivi, riducendo la diversità delle opzioni considerate; da ultimo, la verifica critica che consta della comprensione della capacità del sistema di abilitare o inibire l'abilità degli utenti di

[2] Si vedano, in particolare, *Posture Awareness, Deliberate Switching, Transparent Attribution, Evidence Mapping, Cognitive Audit Trails*.
[3] Le applicazioni operative del *framework*, incluso il c.d. Cognitive Sovereignty Index, sono sviluppate nell'ambito della collaborazione tra Pollicino AIdvisory e ConnectED Mind.

sottoporre a, appunto, una verifica critica le proprie assunzioni e quelle del sistema stesso.[4]

Come si premetteva, il *framework* dei Thinking Quadrants costituisce la struttura metodologica centrale del CogIA aziendale. Articolato nelle quattro posture decisionali — esplorativa (sensemaking, model-led), generativa (analitica, model-led), critica (analitica, human-led) e operativa (sensemaking, human-led) — il *framework* consente di valutare sistematicamente come un sistema di IA interviene su ciascuna fase del processo decisionale, identificando le condizioni in cui il sistema tende a cortocircuitare il *sense-making* umano (bypass diagonale) e i *pattern* di *posture drift* che emergono sotto pressione. In questo senso, l'indagine empirica condotta alla base del framework ha rilevato che il 66% dei dirigenti si concentra nell'Operating Posture, solo il 3% adotta la Critical Posture come default, e il 45% ha sperimentato *output* convincenti ma errati che avrebbero potuto causare danni materiali.

2. Architettura delle salvaguardie cognitive

Il secondo componente riguarda la progettazione delle condizioni strutturali che preservano la qualità del ragionamento organizzativo e il capitale semantico che ne è il substrato, in presenza di sistemi di IA pervasivi.

Questo implica l'introduzione di principi di *cognitive safety by design*, ossia la definizione di domini cognitivi riservati alla deliberazione umana non mediata algoritmicamente, e la progettazione di processi istituzionalizzati di contestazione critica delle raccomandazioni algoritmiche, analoghi alle pratiche *di red teaming* e di *devil's advocacy* documentate nella letteratura sul *decision-making* organizzativo (Janis, 1982; Schulz-Hardt et al., 2006).

Come evidenzia la letteratura, la progettazione delle salvaguardie cognitive deve partire dalla mappatura dei livelli di colonizzazione del System 0 identificati da Riva, nel contributo già richiamato. Tale mappatura comprende tre elementi: la configurazione dell'orizzonte informativo, la normalizzazione del plausibile e la produzione di *prior* cognitivi istituzionali. Per ciascuno di questi livelli, occorre individuare delle misure di salvaguardia che devono interrompere il meccanismo di colonizzazione preservando i benefici funzionali del sistema e la tutela dei diritti degli interessati.

3. Formazione alla sovranità cognitiva

Il terzo componente riguarda lo sviluppo delle competenze individuali e collettive necessarie per interagire con le infrastrutture cognitive di IA preservando l'autonomia del ragionamento e il capitale semantico dell'organizzazione.

Questa formazione mira a sviluppare la capacità di mantenere attivo il livello semantico del processo cognitivo, il livello in cui i dati diventano significato, in

[4] Il CogIA si articola produttivamente con la proposta normativa di Giamminola e Pollicino di introdurre questo strumento come requisito giuridico obbligatorio per i sistemi di AI ad alto impatto istituzionale.

presenza di sistemi che tendono strutturalmente a cortocircuitare questo passaggio.[5]

In termini operativi, la formazione alla sovranità cognitiva deve coprire i tre livelli del System 0, del Sistema 1 e del Sistema 2: al livello del System 0, sviluppare la consapevolezza metacognitiva dei frame cognitivi prodotti dai sistemi di AI — in termini gadameriani, rendere visibile la pre-comprensione esogena come tale; al livello del Sistema 1, preservare e potenziare l'expertise e l'intuizione, come risorse cognitive irriducibili; al livello del Sistema 2, sviluppare le competenze di *critical thinking* applicato all'AI.

4. Integrazione con i meccanismi di *oversight*

Il quarto componente riguarda l'integrazione della dimensione cognitiva nei meccanismi istituzionali di *oversight* dei sistemi di IA. Questo implica che i *board* e i comitati di audit includano tra i propri criteri di valutazione indicatori di salute cognitiva organizzativa, e la creazione di funzioni di Cognitive Security Officer con il mandato specifico di monitorare i rischi cognitivi associati all'adozione pervasiva di sistemi di IA. L'integrazione con *l'oversight* è il punto di connessione tra il framework di governance aziendale qui proposto e la dimensione normativa.

III. Percorsi di applicazione del *framework*

Il *framework* di Cognitive AI Governance sviluppato nelle sezioni precedenti ha trovato i suoi primi percorsi di applicazione operativa in due contesti complementari: la consulenza strategica *executive* e il contesto istituzionale della sicurezza cognitiva. I due percorsi illustrano la traducibilità del *framework* in pratiche organizzative concrete e la sua adattabilità a contesti decisionali differenti per scala, complessità e profilo di rischio.

1. Applicazione nella consulenza strategica *executive*

Il *framework* dei Thinking Quadrants è stato sviluppato ed applicato operativamente nell'ambito dell'*advisory* strategico con dirigenti e board di aziende europee nei settori farmaceutico, beni di consumo e servizi professionali. L'approccio si fonda sul *fieldwork* di 18 mesi con oltre cento dirigenti che ha generato le evidenze empiriche alla base del framework Systemic Zero.

In questo contesto, il *framework* opera a livello di singola organizzazione e di singolo decisore. I dirigenti vengono formati al riconoscimento delle posture cognitive e guidati attraverso processi decisionali strutturati che separano le fasi di esplorazione, generazione, verifica critica e operazionalizzazione. L'obiettivo è costruire un dossier decisionale in cui il processo di ragionamento — e la distribuzione dell'autorità cognitiva tra umano e macchina — sia esplicito e tracciabile. È stata inoltre sviluppata una piattaforma digitale di supporto

[5] Si articola con la tradizione della critical thinking education (Paul e Elder, 2001; Ennis, 1996) e con la ricerca sulla metacognizione (Flavell, 1979; Schraw e Dennison, 1994).

decisionale che implementa i protocolli di *posture switching* e genera *cognitive audit trails* in tempo reale.
I risultati operativi confermano le evidenze del *fieldwork*: i dirigenti che adottano il *framework* riportano una maggiore consapevolezza dei propri *pattern* di interazione con l'AI, una riduzione del *posture drift* sotto pressione e una maggiore capacità di distinguere tra validazione e deliberazione autentica.

2. Applicazione nel contesto istituzionale (ConnectED Mind)

ConnectED Mind, startup innovativa nel settore della sicurezza cognitiva costituita nell'ambito dell'accordo quadro tra il Consiglio Nazionale delle Ricerche (CNR), Fondazione AISES e SPES Academy "Carlo Azeglio Ciampi", rappresenta il secondo percorso di applicazione del *framework*, orientato al contesto istituzionale e alla sicurezza cognitiva su scala organizzativa e di sistema.
L'origine istituzionale di ConnectED Mind, alla convergenza tra ricerca scientifica pubblica, educazione strategica e governance delle politiche economiche e sociali, ne orienta il posizionamento verso la formazione dei decisori pubblici e privati e l'interfaccia con il quadro normativo europeo sulla governance dell'IA. L'offerta si articola in tre linee di difesa integrate: formazione avanzata alla sovranità cognitiva, soluzioni tecnologiche di monitoraggio, e *advisory* strategico.
La dimensione istituzionale di ConnectED Mind aggiunge al *framework* una componente che la consulenza executive non copre: l'applicazione dei principi della cognitive security — già elaborati nella letteratura sulla *cognitive warfare* (Catena, Ditrych e Kovalčíková, 2025) e sulle FIMI (EEAS, 2023) — alla protezione delle architetture decisionali aziendali da minacce cognitive esogene, oltre che dai rischi endogeni analizzati nelle sezioni precedenti.

3. Il Cognitive Sovereignty Index: uno strumento di misurazione congiunto

La convergenza dei due percorsi applicativi ha prodotto il Cognitive Sovereignty Index (CSI): uno strumento di valutazione della postura cognitiva organizzativa che concretizza il Cognitive Impact Assessment proposto da Giamminola e Pollicino (2025) in una metrica misurabile e monitorabile nel tempo. Il CSI è sviluppato congiuntamente da Pollicino AIdvisory (metodologia dei Thinking Quadrants e architettura del CogIA) e ConnectED Mind (applicazione nel contesto istituzionale e di mercato).
Le quattro dimensioni di valutazione del CSI rispecchiano le categorie di rischio identificate nella letteratura interdisciplinare e nel *fieldwork* Systemic Zero: la dipendenza cognitiva (il *cognitive lock-out* elaborato da Giamminola e Pollicino; *l'automation bias* di Parasuraman e Riley); il *framing effect* (System 0 thinking di Chiriatti et al.; convergenza narrativa dei LLM; Narrative Lock-In identificato nel framework dei Thinking Quadrants); la capacità residua di verifica critica (Critical Posture dei Thinking Quadrants; critical thinking education); e l'esposizione a manipolazioni cognitive esterne (cognitive warfare; FIMI).

Il CSI non misura ciò che è già visibile, le performance tecniche dei sistemi algoritmici, la loro conformità normativa, i loro output, ma porta alla luce ciò che opera nel dominio pre-decisionale: il grado di autonomia cognitiva dell'organizzazione, la distribuzione dell'autorità epistemica tra umani e macchine, e la capacità residua di deliberazione critica indipendente. L'*output* del processo è un Cognitive Health Index che consente alle organizzazioni di monitorare la propria sovranità cognitiva nel tempo e di rispondere ai requisiti del CogIA nella misura in cui questo strumento venga istituzionalizzato nel quadro normativo europeo.

IV. Conclusioni

Il presente saggio ha argomentato che la *governance* aziendale dell'intelligenza artificiale si trova di fronte a una sfida strutturale che i *framework* attualmente disponibili non sono attrezzati ad affrontare adeguatamente: governare il potere cognitivo che i sistemi di IA esercitano sulle architetture decisionali delle organizzazioni, prima e indipendentemente dai rischi tecnici, etici e reputazionali associati ai loro output visibili.

Il fondamento filosofico di questa sfida è duplice e convergente. Il principio *visibilia ex invisibilibus* — il visibile nasce dall'invisibile — rivela che ogni decisione visibile affonda le radici in strutture invisibili che la costituiscono prima che essa possa essere deliberata. Si tratta di strutture che i sistemi di AI generativa imparano a configurare con crescente sofisticazione attraverso la colonizzazione del System 0 documentata dalla letteratura qui richiamata.

In merito alla sostenibilità e alla traducibilità operativa dell'approccio, è, infine, dimostrata dai primi percorsi di applicazione del *framework*, nella consulenza strategica *executive* attraverso Pollicino AIdvisory e nel contesto istituzionale della sicurezza cognitiva attraverso il progetto ConnectED Mind. Da ultimo, il Cognitive Sovereignty Index, sviluppato congiuntamente, rappresenta un primo strumento concreto per tradurre la proposta normativa del CogIA in metrica misurabile e governabile, aprendo la strada verso quella standardizzazione che il quadro normativo europeo richiederà come passo successivo.

La posta in gioco della mente umana come campo di battaglia è rappresentata dal capitale semantico, inteso come risorsa cognitiva irriducibile che distingue il giudizio umano dalla computazione algoritmica e che costituisce il fondamento di ogni controllo umano genuino sull'AI.

La proposta si sostanzia in un fondamento normativo che poggia le sue basi nella distinzione tra *human oversight* formale e *human oversight* sostanziale, la proposta del *Cognitive Impact Assessment* come strumento giuridico obbligatorio, la categoria del *cognitive lock-out* come danno giuridicamente rilevante, e la tesi della sovranità cognitiva come diritto fondamentale "costituzionalizzabile" nel quadro della Carta dei Diritti Fondamentali dell'Unione europea.

POLLICINO & PARTNERS
AIDVISORY

GLOSSARIO DEI CONCETTI CHIAVE

System 0: Livello cognitivo pre-decisionale e non umano, introdotto da Chiriatti, Ganapini, Panai, Ubiali e Riva (Nature Human Behaviour, 2024), che opera al di sotto della soglia della consapevolezza e precede strutturalmente i System 1 (intuitivo) e System 2 (riflessivo) di Kahneman. Risiede nell'infrastruttura algoritmica che filtra, ordina e pre-seleziona le informazioni prima che raggiungano la mente consapevole.

Cognitive Infrastructure Studies (CIS): Dominio interdisciplinare introdotto da Riva che riconcettualizza i sistemi IA come infrastrutture cognitive, sintetizzando Science and Technology Studies, scienze cognitive, sociologia digitale, infrastructure theory e teoria argomentativa.

Cognitive Governance: Paradigma di governance elaborato da Giamminola e Pollicino che va oltre la compliance regolatoria per intervenire ex ante sulle architetture decisionali e sui meccanismi di sensemaking strutturati dall'IA.

Intelligent Choice Architectures (ICA): Concetto del MIT Sloan Management Review (Schrage & Kiron, 2025) che descrive sistemi IA generativi e predittivi capaci di ridisegnare dinamicamente l'insieme delle opzioni percepite come rilevanti e accessibili dai decisori umani.

Cognitive Impact Assessment (CogIA): Strumento regolatorio proposto da Giamminola e Pollicino, complementare all'AI Act, per valutare l'impatto dei sistemi IA sull'autonomia cognitiva e sui processi decisionali umani.

Thinking Quadrants: Framework metodologico sviluppato da Giamminola nell'ambito della ricerca Systemic Zero, che articola il CogIA attraverso quattro posture decisionali (Esplorativa, Generativa, Critica, Operativa) lungo due assi: modalità cognitiva e strategic ownership.

Cognitive Lock-Out: Rischio identificato nel framework Systemic Zero: lo spostamento progressivo e irreversibile della strutturazione del pensiero e delle priorità decisionali verso sistemi IA esterni e opachi, con riduzione della capacità di deliberazione cognitiva indipendente.

Cognitive Sovereignty Index (CSI): Strumento di valutazione della postura cognitiva organizzativa sviluppato congiuntamente da Pollicino AIdvisory e ConnectED Mind, che operazionalizza il CogIA in una metrica misurabile e monitorabile nel tempo attraverso quattro dimensioni: dipendenza cognitiva, framing effect, capacità di verifica critica, esposizione a manipolazioni esterne.

Cognitive Health Index: Output del processo CSI: metrica aggregata che consente alle organizzazioni di monitorare la propria sovranità cognitiva nel tempo e di rispondere ai requisiti emergenti del CogIA nel quadro normativo europeo.

Adaptive Invisibility: Caratteristica distintiva delle infrastrutture cognitive per cui, più i sistemi IA diventano sofisticati e integrati, più la loro influenza si fonde seamlessly nei flussi cognitivi, rendendo progressivamente più difficile rilevare o resistere alla mediazione algoritmica.

Infrastructure Breakdown Methodologies: Approcci metodologici proposti da CIS che rendono visibile il ruolo cognitivo invisibile dei sistemi IA simulando condizioni di ritiro, degradazione o interruzione algoritmica dopo periodi di abituazione, rivelando le dipendenze cognitive attraverso il meccanismo del guasto.

FIMI (Foreign Information Manipulation and Interference): Categoria analitica dell'EEAS (2023) che descrive operazioni coordinate di manipolazione informativa di origine straniera; nel contesto di ConnectED Mind, indica il vettore esogeno di minacce alle architetture decisionali organizzative che il Cognitive Sovereignty Index è chiamato a misurare e monitorare.