

L'INTELLIGENZA ARTIFICIALE NELL'AMBITO DELL'ATTIVITA' D'IMPRESA

***L'esperienza della Compliance di Intesa
Sanpaolo***

Torino, 17 luglio 2025

Valerio Cencig

Digital for Compliance

DIGITAL FOR COMPLIANCE

I progetti di I.A. in ambito Compliance



Ambito	Fornitore	Stato	Descrizione
REGULATORY COMPLIANCE	▪ eComm Surveillance	▪ Nice (Shield)	▪ Sorveglianza conversazioni vocali e scritte (richiesta NFA per Dodd-Frank Act)
	▪ Regulatory Monitorig	▪ Aptus.AI (Lisa)	▪ Analisi nuova normativa e relativi contenuti
	▪ Regulatory Chatbot (Privacy)	▪ LexGO	▪ Chatbot per rispondere ai colleghi di filiale sulla Normativa Privacy
	▪ Access Control (Privacy)	▪ Custom	▪ Controllo accessi non autorizzati a dati personali
ANTI FINANCIAL CRIME	▪ 4eyes control detection Embarghi	▪ Pelican	▪ Controlli 4eyes sulle detection generate dai sistemi di controllo Embarghi
	▪ Negative News Screening	▪ Expert.AI	▪ Controllo nel continuo di adverse media, a supporto dei processi di onboarding, periodic/event driven review
	▪ Due Dlligence	▪ GPT-4 Turbo	▪ Utilizzo di generative AI per comporre la Nota Parere e di modelli di AI per differenziare percorsi di KYC risk based

DIGITAL FOR COMPLIANCE

L'Anti Financial Crime Digital Hub



Mission

- L'AFC Digital Hub, oggi pienamente operativo, è una società consortile nata nel Gruppo Intesa Sanpaolo con l'obiettivo di **contrastare i crimini finanziari attraverso l'impiego delle nuove tecnologie e dell'Intelligenza artificiale**
- Si fonda sui **2 principi di Innovazione e Cooperazione** (con partner scientifici e istituzioni finanziarie), ponendosi la finalità di incrementare l'efficacia dei presidi della Banca nel contrasto al crimine finanziario e, in prospettiva, di contribuire a quella del sistema nazionale ed europeo

Informazioni societarie

Soci

INTESA SANPAOLO

INTESA SANPAOLO
INNOVATION CENTER



UNIVERSITÀ
DI TORINO



CENTAI



Politecnico
di Torino



Sede: Torino, Grattacielo ISP



Data di Costituzione: 16 giugno 2022

DIGITAL FOR COMPLIANCE

Le «challenge» dell'AFC Digital Hub



	Challenge	Partner	Obiettivo
Funzionali	Velocity	ISI Foundation	Evoluzione delle tecniche di transaction monitoring, attualmente basate su algoritmi con accumulatori e soglie di alerting per la ricerca del red flag di "Velocity" , usando tecniche di analisi dei grafi temporali
	Reduce Expected Detections	INTESA SANPAOLO Data Office	Sviluppo di algoritmi di machine learning per ottimizzare l'effort dedicato dagli analisti del Competence Center di transaction monitoring alla valutazione degli alert generati dagli applicativi di monitoraggio (Scenario «Carte Prepagate – Persone fisiche» Contante e Traffico Anomalo) per il perimetro delle persone fisiche)
	Hidden Risk	Banca d'Italia	Identificazione univoca delle controparti (c.d. entity resolution) ed identificazione di potenziali scenari di anomalie (c.d. anomaly detection), non note a priori
	Improving Efficiency of Detection Analysis	INTESA SANPAOLO Data Office	Sviluppo di un modello per l' analisi e la valutazione degli alert generati dagli applicativi di monitoraggio, che ne abiliti un trattamento risk-based differenziato anche tramite l'analisi degli aspetti soggettivi
	Connections & Channeling	Banca d'Italia	Analisi dell' evoluzione nel tempo delle reti di relazione costruite attraverso dati transazionali e legami anagrafici al fine di identificare comportamenti potenzialmente sospetti dei clienti
	Multicriteria Anomaly Detection	Banca d'Italia	Implementazione di una porzione di sistema di transaction monitoring in grado di identificare focal entity e relative transazioni anomale con tassi di efficienza ed efficacia migliori rispetto a quelli dei sistemi attuali (focus su perimetro «movimentazione di conti correnti»)
Abilitanti	Federated Learning	Banca d'Italia	Studio e sperimentazione di una metodologia che abilita un approccio collaborativo tra banche nell'addestramento di modelli di machine learning, senza prevedere uno scambio di informazioni tra i soggetti collaboranti e quindi nel rispetto dei vincoli Privacy
	Explainable AI	Geometric logo	Sviluppo di un prototipo in grado di identificare, con approccio qualitativo, l' explainer più adeguato in funzione degli algoritmi e dei dati utilizzati
	Dati Sintetici	Geometric logo	Sviluppo di una metodologia e di un tool per la valutazione della qualità dei dati sintetici generati a partire da un dataset di dati reali
	Quantitative Explainable AI	Geometric logo	Evoluzione del prototipo per l'identificazione degli explainer, definendo e sviluppando un modello quantitativo di classificazione degli explainer sulla base delle metriche individuate: effective compactness, fidelity e stability
	Explainable AI System	Geometric logo	Sviluppo di un framework utilizzabile e richiamabile da un generico applicativo che faccia uso di ML/AI da utilizzare: i) ex-ante per individuare i migliori explainer disponibili per spiegare le predizioni del modello di ML individuato e del set di dati utilizzati; ii) ex-post per la spiegazione delle predizioni del modello tramite traduzione in linguaggio naturale

DIGITAL FOR COMPLIANCE

Le sfide poste dall'intelligenza artificiale



Sfida	Descrizione	Soluzioni
Dataset sbilanciati	Le transazioni sospette, sul totale, sono molto poche ("ago nel pagliaio")	Dati sintetici per "amplificare" il segnale
Assenza di una ground truth	Attualmente anche i modelli supervisionati sono basati su una "presunzione" di Illecito (presenza di una SOS) non confermato	Feedback da parte delle Autorità
Privacy	L'addestramento dei modelli di intelligenza artificiale richiede di poter lavorare su dati granulari (transazione/cliente)	Tecniche di anonimizzazione: PET, Federated Learning, ecc.
Spiegabilità degli algoritmi	Molte tecniche di intelligenza artificiale non sono nativamente spiegabili ("blackbox")	Explainer
Complessità degli algoritmi	Lo sviluppo e la manutenzione di sistemi di intelligenza artificiale richiede competenze tecniche e funzionali	GDL multidisciplinari (funzionali e quantitativi)
Data Quality	Le performance dei Sistemi di intelligenza artificiale sono estremamente sensibili alla qualità dei dati usati nel training	Applicazione rigorosa della data governance
Cambio di paradigma	Dalla ricerca di transazioni sospette alla ricerca di schemi comportamentali inusuali Dalla Red Flag Analysis alla Feature Selection Dalla Coverage Analysis a test di efficacia sostanziali	

Compliance for Digital

COMPLIANCE FOR DIGITAL

La governance dell'intelligenza artificiale



Principali Action



Linee guida e regole

Definizione ed aggiornamento nel tempo, anche in relazione all'evoluzione tecnologica, di:

- **Linee Guida**
- **Regole**
- **Metriche e Soglie**



Processi & Controlli

Disegno, implementazione ed ottimizzazione continua dei **processi end to end** di **acquisizione/sviluppo** di un Sistema di I.A. e dei controlli necessari per garantire una **conformity by design** ed un **monitoraggio nel continuo**, attraverso la definizione di:

- Guide di Processo
- Obiettivi di controllo di 1° e 2° livello



Tool & Abilitatori

Sviluppo di **strumenti** finalizzati a supportare i **processi di governance** e **presidio dei rischi**, come ad es.:

- **Repository** centralizzato dei Sistemi di I.A.
- Strumenti di spiegabilità (**Explainer**)
- Strumenti di **interceettazione, controllo e rimozione dei bias** presenti sui dati o negli algoritmi
- **Dashboard** dei controlli



Diffusione della cultura

- Sessioni di **Induction al Board** ed al Management aziendale attraverso momenti di informazione specifica
- Programmi di **formazione differenziati in funzione del ruolo/mestiere aziendale**, con profondità e contenuti progettati in funzione delle rispettive esigenze

L'implementazione di una solida **Governance dell'I.A.** è un **elemento chiave** per creare **fiducia nell'utilizzo** dell'Intelligenza Artificiale e quindi per sfruttarne in modo sicuro e responsabile i potenziali benefici..

È necessario però prevedere una **Governance «dinamica»**, capace di riflettere la **continua e rapida evoluzione tecnologica**.

COMPLIANCE FOR DIGITAL

I rischi dell'intelligenza artificiale ed i relativi presidi

Rischi già noti connessi alla tecnologia

Come tutti i sistemi software i Sistemi di I.A. **presentano rischi correlati alla (cyber)sicurezza, data quality, robustezza, data privacy**, ecc. Tali rischi sono presidiati sulla base di Norme esterne ed interne al Gruppo **già esistenti**

Rischi nuovi propri dell'I.A.

I Sistemi di I.A. presentano anche **rischi specifici**, collegati alla natura di «apprendimento automatico», quali ad esempio:

- la presenza di discriminazioni involontarie (**Fairness**)
- la mancanza di informazioni esplicite sul funzionamento o la difficoltà nel comprendere i risultati prodotti dal Sistema di I.A. (**Trasparenza e Spiegabilità**)
- la mancanza di controllo (**Supervisione Umana**)

Rischi aggiuntivi collegati all'I.A. generativa

I Sistemi di I.A. generativa sono inoltre generalmente associati a **rischi peculiari** rispetto agli altri Sistemi di I.A. tra cui:

- rischi legati alla generazione di risposte sbagliate ma «credibili» (**Allucinazioni**)
- rischi di violazione della proprietà intellettuale quando addestrate su dataset contenenti contenuti protetti da copyright e licenze (**Proprietà intellettuale**)

Rischi connessi ai Sistemi di I.A.		
RISCHI	ORIGINE	PRESIDI DA APPLICARE
Discriminazioni involontarie	Possibilità di trasmettere, amplificare o introdurre involontariamente bias discriminatori in un Sistema di I.A.	Misure di Fairness
Mancata comprensione	Il funzionamento dell'algoritmo avviene sulla base di correlazioni auto-apprese e non sulla base di regole definite dall'uomo	Misure di Trasparenza e Spiegabilità
Perdita di controllo	Tendenza alla eccessiva fiducia ed alla delega nei confronti dei Sistemi di I.A., presunti superiori e più oggettivi nei risultati	Misure di Sorveglianza Umana

Principali* Rischi aggiuntivi collegati ai Sistemi di I.A. generativa		
RISCHI	ORIGINE	PRESIDI DA APPLICARE
Allucinazioni	Rischio di generare risposte sbagliate ma presentate in modo tale da renderne difficile il riconoscimento	Misure di Sorveglianza Umana, Guardrail, Grounding, Prompting
Proprietà Intellettuale	Rischio di violazione della proprietà intellettuale se il Sistema di I.A. è stato addestrato includendo contenuti protetti da copyright e licenze	Verifica dei dataset di addestramento

(*) Altri rischi: diffamazione, instabilità delle risposte, violazione della confidenzialità

COMPLIANCE FOR DIGITAL

I sistemi di intelligenza artificiale in Banca

Sistemi Vietati

Nel Gruppo Intesa Sanpaolo **non si riscontra l'utilizzo di Sistemi di I.A. vietati.**

Alcuni esempi:

- **Sistemi di scoring creditizio** basati sui dati personali dei clienti, sui loro comportamenti e sulle loro interazioni all'interno della società in cui vivono
- Sistemi di marketing **manipolatori** o che fanno uso di tecniche **subliminali**, per distorcere il comportamento dei consumatori

Sistemi ad Alto rischio

I Sistemi di I.A, potenzialmente ad Alto Rischio sono quelli impiegati nei seguenti ambiti: Credito, HR, Assicurazioni; Sistemi Frodi escluse, AML in discussione.

Alcuni esempi:

- Sistemi di I.A. utilizzati per la **valutazione dell'affidabilità creditizia** delle persone fisiche e per la determinazione dei premi assicurativi per polizze vita o salute
- Sistemi di I.A **impiegati per la selezione di candidati** da assumere o per il monitoraggio delle prestazioni dei lavoratori

Sistemi Non ad Alto Rischio

La maggior parte dei Sistemi di I.A. usati nel Gruppo rientrano nella categoria non ad Alto Rischio.

Alcuni esempi:

- Sistemi per **ottimizzare i percorsi e la frequenza** del caricamento contante negli ATM
- Sistemi per la **valutazione** delle **società** quotate che, sulla base del bilancio e del mercato di appartenenza, attribuiscono un rating alla società analizzata

Tali Sistemi di I.A. possono essere sia ad Alto rischio che non ad Alto rischio

Sistemi con Obbligo di informazione

Rientrano in tale categoria i Sistemi di I.A. che interagiscono con i Clienti o i Dipendenti. Ad esempio:

- **Assistenti virtuali** (es. chatbot), utilizzati per fornire assistenza e rispondere alle domande testuali o vocali dei clienti/utenti relativamente al catalogo prodotti e servizi disponibile



www.odcec.torino.it

www.linkedin.com/company/odcec-torino/

www.youtube.com/channel/UCBUHnLEOEHA6YY-MLr8vG8A/videos