Università
Ca'Foscari
Venezia

Corso di Laurea magistrale

in Economia e Finanza

Tesi di Laurea

Antiriciclaggio: l'impatto dell'intelligenza artificiale sul monitoraggio delle transazioni

Relatore
Ch. Prof. Andrea Minto

Laureando
Riccardo
Tanci
Matricola
889801

Anno Accademico
2025 / 2026

Sommario

Introduzione ***4***

Capitolo 1 - Il quadro normativo e operativo delle Segnalazioni di Operazioni Sospette ***7***

1.1 L'obbligo di segnalazione attiva nel D.Lgs. 231/2007 e il ruolo di UIF e GdF . 7

1.1.1 Genesi e architettura del D.Lgs. 231/2007 7
1.1.2 Il Pacchetto AML 2024: verso un sistema europeo armonizzato 8
1.1.3 I soggetti obbligati e la platea delle "sentinelle" 10
1.1.4 L'obbligo di segnalazione: natura giuridica e presupposti 10
1.1.5 Il ruolo istituzionale della UIF e della Guardia di Finanza 11

1.2 Il processo attuale di individuazione dell'anomalia: monitoraggio automatico e valutazione discrezionale 13

1.2.1 La cornice normativa del processo di detection 13
1.2.2 La procedura GIANOS e l'architettura del monitoraggio automatico 13
1.2.3 Dall'alert alla valutazione: il problema dei falsi positivi 14
1.2.4 L'evoluzione verso logiche predittive 15
1.2.5 La rilevazione umana: il presidio insostituibile della rete 15

1.3 Le criticità odierne: il fenomeno del Defensive Reporting e l'elevato numero di falsi positivi. 17

1.3.1 Il sovraccarico informativo come patologia del sistema 17
1.3.2 Il *Defensive Reporting*: genesi e razionalità economica 17
1.3.3 L'inefficienza economica e il paradosso dei falsi positivi 19
1.3.4 La carenza di *feedback* e l'impossibilità di taratura 19
1.3.5 L'andamento recente delle segnalazioni 20

1.4 Le nuove Istruzioni UIF del 18 dicembre 2025: la disciplina rinnovata della collaborazione attiva 22

1.4.1 Contesto e ratio del nuovo Provvedimento 22
1.4.2 Il rifiuto degli automatismi come risposta al *Defensive Reporting* 23
1.4.3 Il processo valutativo in tre fasi e la tracciabilità della decisione 24
1.4.4 L'apertura all'intelligenza artificiale e la centralità dell'intervento umano 24
1.4.5 Il rafforzamento del flusso di ritorno 25
1.4.6 Conclusioni del quadro normativo 25

Capitolo 2 - L'evoluzione dei modelli di controllo: dall'approccio Rule-Based ai sistemi predittivi ***27***

2.1 Limiti degli schemi rigidi dei sistemi tradizionali 27

2.2 Le nuove frontiere del *Transaction Monitoring*: l'uso dell'Intelligenza Artificiale per l'identificazione di anomalie complesse 29

2.2.1 Apprendimento supervisionato e non supervisionato 30
2.2.2 I modelli principali: alberi decisionali, foreste casuali, *gradient boosting* e reti neurali 31
2.2.3 L'analisi delle reti: dalla singola operazione alla struttura relazionale 34
2.2.4 I risultati concreti: benefici, diffusione effettiva e limiti 36
2.2.5 La soglia di segnalazione: ciò che la tecnologia non cambia 39
2.2.6 Caso studio della UIF: un modello di *machine learning* per l'identificazione di aziende collegate alla criminalità organizzata in Italia 41

2.3 L'impatto organizzativo: come cambia il lavoro del delegato SOS con il supporto di strumenti tecnologici evoluti (relazione *System-to-Human*) 43

2.3.1 Dal filtro degli *alert* all'analisi qualificata: la trasformazione del ruolo 43

2.3.2 La relazione *System-to-Human*: automazione, giudizio e responsabilità 44

2.4 Verso il problema della "black box": efficacia tecnica e opacità decisionale ..47

Capitolo 3 - Profili di governance e il problema della "Black Box" 50

3.1 La necessità di motivare la segnalazione: il requisito della "spiegabilità" delle decisioni algoritmiche 50

3.1.1 L'obbligo di motivare il sospetto: il fondamento normativo 51

3.1.2 La spiegabilità tecnica: cosa può e cosa non può offrire l'intelligenza artificiale..... 53

3.2 La responsabilità dell'intermediario e del delegato SOS.....58

3.2.1 Il quadro sanzionatorio dell'antiriciclaggio e la sua struttura..... 58

3.2.2 Il delegato SOS tra automazione e responsabilità personale 59

3.2.3 La responsabilità dell'intermediario e la scelta del sistema 60

3.2.4 Il fornitore fuori dal perimetro: il nodo dell'eterodirezione..... 60

3.2.5 Un problema di allocazione, non soltanto di imputazione 62

3.3 Riflessioni sull'AI Act.....63

3.3.1 L'architettura dell'AI Act e il metodo del rischio..... 63

3.3.2 Il monitoraggio delle transazioni, in ambito AML, è un sistema ad alto rischio? 64

3.3.3 Il diritto alla spiegazione e la sua inapplicabilità al segnalato 65

3.3.4 Il rinvio degli obblighi e lo stato attuale 68

3.3.5 Un paradosso e le sue implicazioni..... 68

Conclusioni..... 70

Bibliografia..... 74

Monografie e contributi dottrinali 74

Articoli e contributi in riviste e atti di convegno 74

Documenti istituzionali 76

Normativa di riferimento..... 76

Giurisprudenza 78

Sitografia 79

Introduzione

Il riciclaggio di denaro rappresenta oggi una minaccia strutturale per l'integrità dei mercati finanziari globali. Quando le organizzazioni criminali riescono a mascherare l'origine illecita di ingenti capitali e a reimmetterli nell'economia legale attraverso operazioni in apparenza regolari, il danno non si limita alla distorsione della concorrenza. A essere colpite sono anche, e più in profondità, la stabilità delle istituzioni democratiche e la fiducia dei cittadini nel sistema finanziario[1]. In risposta a tale minaccia strutturale, l'ordinamento giuridico nazionale ed europeo si è attivato per una trasformazione paradigmatica: si è transitati da un approccio meramente repressivo e reattivo (incentrato sul sanzionamento e sulla persecuzione penale ex-post del riciclatore) a una strategia preventivo-amministrativa fondata sulla collaborazione attiva e sistematica del settore privato con le autorità pubbliche[2].

In tale prospettiva, il sistema italiano di prevenzione del riciclaggio, fondato sul Decreto Legislativo n. 231 del 21 novembre 2007, si basa su un principio cardine: gli intermediari finanziari, i professionisti e, più in generale, i soggetti che operano a presidio dei flussi economici non sono meri spettatori passivi, ma sono chiamati ad assumere un ruolo attivo di "sentinelle" del sistema. Tra gli strumenti operativi che tale sistema mette a disposizione, la Segnalazione di Operazione Sospetta (di seguito, SOS) rappresenta il fulcro dell'intero impianto preventivo: essa costituisce il punto di raccordo necessario e istituzionale tra l'operatore privato e le autorità di controllo.

La presente tesi si propone di analizzare la SOS nella sua complessità multidimensionale (normativa, operativa e tecnologica) esponendo i limiti e le criticità degli attuali sistemi di monitoraggio delle transazioni ed analizzando l'impatto in termini di efficienza che l'utilizzo di sistemi predittivi possa portare in tale ambito. Ci si chiede quindi, in un'epoca in cui l'intelligenza artificiale promette di rivoluzionare i processi decisionali, quale sia e quale debba essere il ruolo del fattore umano nell'ambito delle segnalazioni di operazioni sospette.

Il tema acquista una rilevanza crescente alla luce dei dati più recenti. Le segnalazioni pervenute all'Unità di Informazione Finanziaria per l'Italia (UIF) hanno superato nel 2025 le 160.000 unità annue: un volume che riflette l'ampliamento della platea dei soggetti obbligati e l'emergere dei nuovi fenomeni criminali, ma anche, come si avrà modo di argomentare, una tendenza patologica alla c.d. *defensive reporting*, ossia alla segnalazione difensiva a scopo cautelare[3]. Una mole informativa così ingente rischia paradossalmente di depotenziare l'efficacia dell'azione investigativa, disperdendo le risorse analitiche in un mare di "falsi positivi" e rendendo più ardua l'individuazione delle reali operazioni illecite.

---

[1] Sulla dimensione del fenomeno e sulla sua portata sistemica cfr. LANNOO K. – PARLOUR R., *Anti-Money Laundering in the EU: Time to get serious*, CEPS-ECRI Task Force Report, Bruxelles, 2021.

[2] Sul passaggio dal modello repressivo a quello preventivo-amministrativo si veda RAZZANTE Ranieri, *Manuale di legislazione e prassi dell'antiriciclaggio,* Giappichelli, Torino, 2023.

[3] Cfr. UIF – Unità di Informazione Finanziaria per l'Italia, *Rapporto annuale per il 2025*, n. 18, 2026 disponibile su: uif.bancaditalia.it.

Il filo conduttore della presente ricerca si articola attorno alla tesi che l'intelligenza artificiale non costituisca, né possa costituire, un sostituto del giudizio umano nell'ambito delle SOS, bensì uno strumento di supporto capace di ottimizzare il processo, ridurre i falsi positivi e ampliare la capacità di rilevare anomalie complesse. La decisione finale, e con essa la responsabilità giuridica, rimangono totalmente in capo al decisore umano — il delegato SOS o il responsabile antiriciclaggio — in quanto presidio insostituibile di valutazione qualitativa e di imputabilità giuridica.

La trattazione si sviluppa in tre capitoli. Il Capitolo 1 ricostruisce il quadro normativo e operativo delle SOS, analizzando gli obblighi previsti dall'art. 35 del D.Lgs. 231/2007, il ruolo istituzionale della UIF e della Guardia di Finanza, e le criticità strutturali del sistema attuale, con particolare riguardo al fenomeno del *defensive reporting* e al problema dei falsi positivi. Il capitolo dà conto anche del pacchetto AML europeo e delle nuove istruzioni della UIF. Il Capitolo 2 indaga l'evoluzione dei modelli di controllo, dai sistemi *rule-based* tradizionali — come la procedura GIANOS — alle nuove frontiere rappresentate dall'intelligenza artificiale e dal machine learning applicati al *transaction monitoring*; si esamina inoltre l'impatto organizzativo di tali innovazioni sulla figura del delegato SOS nella relazione system-to-human. Il Capitolo 3 affronta i profili di governance e la questione della c.d. "black box", analizzando il requisito di spiegabilità delle decisioni algoritmiche, la responsabilità giuridica dell'intermediario e gli orientamenti normativi emergenti, con riferimento all'AI Act europeo e alle indicazioni di Banca d'Italia e UIF.

La metodologia adottata è di natura giuridico-economica, combinando l'analisi della normativa primaria e secondaria con l'esame della documentazione istituzionale e della dottrina specialistica. Il momento storico in cui la presente ricerca si colloca ne accresce ulteriormente la rilevanza: il quadro normativo di riferimento a livello europeo è in profonda trasformazione su due fronti convergenti. Sul versante antiriciclaggio, il 19 giugno 2024 è stato pubblicato nella Gazzetta Ufficiale dell'Unione Europea l'AML Package, che si compone del Regolamento Antiriciclaggio (UE) 2024/1624 — c.d. *single rulebook* —, della VI Direttiva Antiriciclaggio (UE) 2024/1640 e del Regolamento (UE) 2024/1620 istitutivo della nuova Autorità europea per la lotta al riciclaggio e al finanziamento del terrorismo (AMLA), con sede a Francoforte[4]. L'AMLA ha avviato le proprie attività operative dal 1° luglio 2025 e il *single rulebook* diverrà pienamente applicabile a decorrere dal 10 luglio 2027: si tratta di una riforma che mira a superare la frammentazione normativa nazionale attraverso regole direttamente applicabili e uniformi in tutti gli Stati membri, ridisegnando in profondità il perimetro dei soggetti obbligati e i meccanismi di supervisione. Sul versante dell'intelligenza artificiale, il Regolamento (UE) 2024/1689, noto come AI Act, è entrato in vigore il 1° agosto 2024 e costituisce il primo quadro giuridico organico al mondo in materia. Esso gradua gli obblighi in funzione del rischio, riservando ai sistemi classificati ad alto rischio requisiti stringenti di trasparenza e di sorveglianza umana; il relativo calendario applicativo è stato peraltro differito nel 2026, a fronte dei ritardi nella predisposizione degli strumenti tecnici necessari. Come si argomenterà nel terzo capitolo, tuttavia, i sistemi di monitoraggio delle

[4] *AML Package*: Regolamento (UE) 2024/1624 del Parlamento europeo e del Consiglio del 31 maggio 2024; Direttiva (UE) 2024/1640 del 31 maggio 2024 (VI Direttiva Antiriciclaggio); Regolamento (UE) 2024/1620 del 31 maggio 2024 (Regolamento AMLA); pubblicati nella Gazzetta Ufficiale dell'Unione Europea il 19 giugno 2024.

transazioni impiegati dagli intermediari non rientrano fra quelli qualificati ad alto rischio: una circostanza tutt'altro che neutrale, che colloca il settore antiriciclaggio al di fuori del perimetro delle garanzie più intense proprio mentre gliene attribuisce, per altra via, gli oneri[5]. Sul piano nazionale, la stessa Unità di Informazione Finanziaria ha emanato nuove istruzioni sulle segnalazioni di operazioni sospette, in vigore dal 1° luglio 2026, che per la prima volta si occupano anche dell'impiego dell'intelligenza artificiale: se ne tratterà nel primo capitolo.

L'obiettivo non è meramente descrittivo, ma critico: si intende offrire una valutazione argomentata delle tensioni ancora irrisolte tra innovazione tecnologica, garanzie di imputabilità e centralità del giudizio umano in un dominio — quello dell'antiriciclaggio — in cui la posta in gioco non è soltanto la conformità normativa, ma la stessa integrità del sistema economico-finanziario.

[5] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale, in GUUE L 2024/1689 del 12 luglio 2024, entrato in vigore il 1° agosto 2024. Il termine di applicazione delle disposizioni relative ai sistemi ad alto rischio elencati nell'Allegato III, originariamente fissato al 2 agosto 2026, è stato differito al 2 dicembre 2027 dal Regolamento (UE) 2026/1744 dell'8 luglio 2026 (Omnibus digitale sull'IA), pubblicato nella GUUE, serie L, del 24 luglio 2026 ed entrato in vigore il 27 luglio 2026.

Capitolo 1 - Il quadro normativo e operativo delle Segnalazioni di Operazioni Sospette

1.1 L'obbligo di segnalazione attiva nel D.Lgs. 231/2007 e il ruolo di UIF e GdF

1.1.1 Genesi e architettura del D.Lgs. 231/2007

L'impianto normativo italiano in materia di prevenzione del riciclaggio è il risultato di un lungo e stratificato processo di adeguamento agli standard internazionali elaborati dal Gruppo d'Azione Finanziaria Internazionale (GAFI/FATF) e alle successive Direttive dell'Unione Europea. Il testo fondativo di tale sistema è rappresentato dal Decreto Legislativo n. 231 del 21 novembre 2007, con il quale il legislatore ha recepito la Direttiva 2005/60/CE (c.d. Terza Direttiva Antiriciclaggio), dando vita a un corpus normativo organico destinato a sostituire la frammentata disciplina previgente[6].

La scelta sistematica del legislatore — confermata e rafforzata dalle successive riforme — è stata quella di costruire un sistema di prevenzione fondato non già sulla repressione penale ex post, compito riservato al codice penale, ma su un meccanismo di prevenzione ex ante che coinvolga attivamente gli operatori economici e finanziari nella funzione di rilevazione precoce del rischio di riciclaggio. Il D.Lgs. 231/2007 persegue l'obiettivo di trasformare gli intermediari in "sentinelle" del sistema economico[7], imponendo loro di intercettare i flussi finanziari anomali prima che il capitale illecito si consolidi nell'economia legale.

Il decreto ha subito successivamente due revisioni di portata strutturale sul piano interno. La prima, operata con il D.Lgs. 25 maggio 2017, n. 90, ha recepito la IV Direttiva Antiriciclaggio (2015/849/UE), rafforzando i presidi di trasparenza in materia di titolare effettivo e accentuando l'approccio basato sul rischio. La seconda, attuata con il D.Lgs. 4 ottobre 2019, n. 125, ha recepito la V Direttiva (2018/843/UE), estendendo il perimetro soggettivo degli obblighi, con particolare riferimento alle piattaforme di scambio di valute virtuali, e rafforzando la cooperazione tra le Unità di Informazione Finanziaria degli Stati membri[8].

Il decreto articola il sistema di prevenzione attorno a tre pilastri fondamentali. Il primo è l'adeguata verifica della clientela (*Customer Due Diligence*), disciplinata dagli artt. 17 ss. del decreto, che impone ai soggetti obbligati di identificare il cliente e il titolare effettivo, di comprendere la natura e lo scopo del rapporto o della prestazione e di monitorarne l'evoluzione nel tempo. Il secondo pilastro è la conservazione delle informazioni e dei documenti per un periodo di dieci anni, che garantisce la ricostruibilità dei flussi

[6] Il D.Lgs. 21 novembre 2007, n. 231, pubblicato in G.U. n. 290 del 14 dicembre 2007, ha abrogato il previgente D.Lgs. 20 febbraio 2004, n. 56, che aveva recepito la Seconda Direttiva Antiriciclaggio (2001/97/CE).
[7] RAZZANTE Ranieri, *Manuale di legislazione e prassi dell'antiriciclaggio,* Giappichelli, Torino, 2023.
[8] D.Lgs. 25 maggio 2017, n. 90; D.Lgs. 4 ottobre 2019, n. 125.

finanziari a fini investigativi ai sensi dell'art. 31 del decreto. Il terzo — e per le finalità della presente trattazione più rilevante — è la segnalazione di operazioni sospette, disciplinata dall'art. 35, quale obbligo attivo di comunicare alla UIF le operazioni che, per caratteristiche, entità, natura ovvero per le particolari modalità e circostanze di svolgimento, inducano a ritenere che i fondi possano provenire da attività criminosa[9].

Un elemento cardine del sistema, introdotto dalle riforme più recenti, è l'Approccio Basato sul Rischio (*Risk-Based Approach*, RBA). A differenza del sistema precedente, in cui gli adempimenti erano definiti in modo rigido e formale, il RBA impone ai soggetti obbligati di modulare l'intensità dei controlli in proporzione al rischio effettivo di riciclaggio associato al cliente, al prodotto, al canale distributivo e all'area geografica di operatività. Ne consegue che la valutazione non è più un atto puntuale ma un processo continuo e dinamico, che richiede all'intermediario di aggiornare costantemente il profilo di rischio della clientela in risposta all'evoluzione dell'operatività[10].

1.1.2 Il Pacchetto AML 2024: verso un sistema europeo armonizzato

Il quadro normativo descritto è attualmente oggetto di una riforma di portata storica, destinata a ridisegnare in profondità l'architettura antiriciclaggio europea. Il 19 giugno 2024 è stato pubblicato nella Gazzetta Ufficiale dell'Unione Europea il c.d. AML Package, approvato in via definitiva dal Parlamento Europeo il 24 aprile 2024 e dal Consiglio dell'Unione il 30 maggio 2024. Il pacchetto si compone di tre atti normativi tra loro coordinati, ciascuno con una funzione distinta e complementare[11].

Il primo pilastro è il Regolamento (UE) 2024/1624, comunemente denominato *single rulebook*, che costituisce la novità strutturale più rilevante: per la prima volta nella storia della disciplina antiriciclaggio europea, tutte le norme applicabili ai soggetti obbligati del settore privato vengono unificate in un regolamento direttamente applicabile in tutti gli Stati membri, senza necessità di recepimento nazionale. Viene così superata la frammentazione che aveva caratterizzato il sistema delle Direttive, in cui differenze significative di trasposizione da uno Stato membro all'altro creavano asimmetrie normative sfruttabili dai riciclatori attraverso il c.d. *regulatory arbitrage*. Il *single rulebook* diventerà pienamente applicabile a decorrere dal 10 luglio 2027, con alcune eccezioni settoriali prorogate al 2029[12].

Il secondo pilastro è la VI Direttiva Antiriciclaggio (UE) 2024/1640, che disciplina i meccanismi istituzionali che gli Stati membri devono predisporre a livello nazionale — in particolare l'organizzazione delle autorità competenti e la cooperazione tra le Unità di

---

[9] D.Lgs. 21 novembre 2007, n. 231, artt. 17 ss. (adeguata verifica), art. 31 (conservazione), art. 35 (segnalazione di operazioni sospette).

[10] Sul Risk-Based Approach si veda GAFI, *Guidance for a Risk-Based Approach – The Banking Sector*, ottobre 2014, disponibile su: fatf-gafi.org.

[11] *AML Package*: Regolamento (UE) 2024/1624, Direttiva (UE) 2024/1640 e Regolamento (UE) 2024/1620, tutti del 31 maggio 2024, pubblicati nella Gazzetta Ufficiale dell'Unione Europea del 19 giugno 2024.

[12] Regolamento (UE) 2024/1624, art. 90. La piena applicabilità è fissata al 10 luglio 2027, con proroga al 10 luglio 2029 per agenti calcistici e società calcistiche professionistiche ai sensi dell'art. 3, punto 3), lett. n) e o).

Informazione Finanziaria — e abroga la IV Direttiva (2015/849/UE) a decorrere dal 10 luglio 2027, termine entro cui gli Stati membri sono tenuti al recepimento[13].

Il terzo pilastro, e forse quello di maggiore impatto sistemico nel breve periodo, è il Regolamento (UE) 2024/1620, istitutivo della nuova Autorità europea per la lotta al riciclaggio e al finanziamento del terrorismo (*Anti-Money Laundering Authority*, AMLA), con sede a Francoforte. L'AMLA ha avviato le proprie attività operative a partire dal 1° luglio 2025 e rappresenta un'assoluta novità nell'architettura istituzionale europea: per la prima volta, viene istituita un'autorità sovranazionale dotata di poteri di supervisione diretta sui soggetti finanziari classificati ad alto rischio e attivi in più Stati membri, nonché di poteri di supervisione indiretta e di coordinamento nei confronti delle autorità nazionali competenti e delle Unità di Informazione Finanziaria[14]. In caso di gravi o sistematiche carenze nei sistemi di controllo, l'AMLA è inoltre investita di poteri sanzionatori diretti nei confronti dei soggetti obbligati selezionati, con un regime di ammende che può raggiungere importi significativi[15].

Sul piano dei contenuti sostanziali, il Pacchetto AML 2024 introduce alcune novità di rilievo che incidono direttamente sull'operatività dei soggetti obbligati e sul tema oggetto della presente trattazione. In primo luogo, viene ulteriormente rafforzato e standardizzato il processo di adeguata verifica della clientela, con l'obiettivo di ridurre le difformità applicative tra intermediari e tra giurisdizioni. In secondo luogo, il *single rulebook* amplia significativamente la platea dei soggetti obbligati, estendendo gli obblighi antiriciclaggio a nuove categorie di operatori, tra cui i fornitori di servizi di cripto-attività ai sensi del Regolamento MiCA, le piattaforme di *crowdfunding*, gli agenti calcistici e le società calcistiche professionistiche, nonché i commercianti di beni di lusso e di beni culturali[16]. In terzo luogo, vengono rafforzate le disposizioni in materia di trasparenza della titolarità effettiva, con criteri di individuazione più stringenti e un più ampio accesso ai relativi registri da parte delle autorità competenti.

Per le finalità della presente tesi, è di particolare interesse evidenziare come il Pacchetto AML 2024 non si limiti a intervenire sui contenuti sostanziali degli obblighi, ma ponga anche le basi per una standardizzazione degli strumenti e delle metodologie di rilevazione del rischio a livello europeo, in un contesto in cui la componente tecnologica — e in particolare l'impiego dell'intelligenza artificiale nel *transaction monitoring* — acquista un'importanza crescente. L'AMLA avrà il compito di emanare norme tecniche di regolamentazione (*RTS*) e linee guida che orienteranno anche le scelte tecnologiche degli intermediari.

---

[13] Direttiva (UE) 2024/1640, art. 1 e art. 78 (termine di recepimento: 10 luglio 2027) e art. 77 (abrogazione della Direttiva (UE) 2015/849 a decorrere dalla medesima data).

[14] Regolamento (UE) 2024/1620, art. 108 (avvio operatività: 1° luglio 2025).

[15] Regolamento (UE) 2024/1620, art. 22, che attribuisce all'Autorità il potere di irrogare sanzioni pecuniarie ai soggetti obbligati selezionati che violino, intenzionalmente o per negligenza, gli obblighi direttamente applicabili previsti dal Regolamento (UE) 2024/1624 o dal Regolamento (UE) 2023/1113, ovvero non si conformino alle decisioni vincolanti dell'Autorità; cfr. altresì l'Allegato I al medesimo regolamento, in ordine alla determinazione degli importi.

[16] Regolamento (UE) 2024/1624, art. 3 (soggetti obbligati), con estensione a fornitori di servizi su cripto-attività, piattaforme di *crowdfunding*, agenti calcistici, commercianti di beni di lusso e culturali.

1.1.3 I soggetti obbligati e la platea delle "sentinelle"

L'efficacia dell'intero sistema preventivo poggia sulla collaborazione di una pluralità eterogenea di operatori economici, che il D.Lgs. 231/2007 definisce, all'art. 3, come "soggetti obbligati" — categoria destinata ad ampliarsi ulteriormente con l'entrata in applicazione del *single rulebook*. La scelta del legislatore, nazionale ed europeo, di estendere progressivamente tale categoria risponde alla consapevolezza che il riciclaggio non si esaurisce nei canali bancari tradizionali, ma si insinua in tutte le pieghe del sistema economico.

Tra i soggetti obbligati rientrano, in primo luogo, gli intermediari bancari e finanziari: banche, Poste Italiane S.p.A., istituti di moneta elettronica (IMEL), istituti di pagamento, società di intermediazione mobiliare (SIM) e società di gestione del risparmio (SGR). Si tratta della categoria numericamente più significativa in termini di SOS prodotte, circa il 90% del totale, secondo i dati pubblicati dalla UIF nel Rapporto Annuale 2023[17] .

Rientrano nella platea anche i professionisti — dottori commercialisti, esperti contabili, avvocati e notai — il cui contributo è particolarmente prezioso nelle fasi di costituzione societaria, nelle operazioni immobiliari e nella gestione patrimoniale. Questi soggetti operano, infatti, in contesti in cui l'anomalia si manifesta più frequentemente in forma qualitativa che quantitativa e sfugge pertanto ai sistemi di rilevazione automatica. Sono altresì obbligati gli operatori non finanziari tra cui: commercianti di beni di alto valore, agenti immobiliari e gli operatori di gioco, esposti al rischio di riciclaggio attraverso sistemi di scommesse e vincite fittizie.

Di particolare rilievo sistematico è l'estensione già operata dalla V Direttiva ai prestatori di servizi relativi a valute virtuali (VASP, *Virtual Asset Service Providers*): con il recepimento di tale direttiva, anche le piattaforme di scambio di criptovalute e i fornitori di servizi di portafoglio digitale sono pienamente assoggettati agli obblighi antiriciclaggio, con obbligo di iscrizione nell'apposito registro tenuto dall'OAM (Organismo Agenti e Mediatori)[18]. Con il Pacchetto AML 2024 tale estensione viene ulteriormente rafforzata e coordinata con la disciplina MiCA, garantendo una copertura normativa più coerente e aggiornata rispetto all'evoluzione del mercato delle cripto-attività.

1.1.4 L'obbligo di segnalazione: natura giuridica e presupposti

Il cuore dell'obbligo di collaborazione attiva è contenuto nell'art. 35 del D.Lgs. 231/2007, che costituisce la norma cardine dell'intero sistema SOS. La disposizione stabilisce che i soggetti obbligati trasmettano senza ritardo una segnalazione alla UIF «quando sanno,

[17] UIF – Unità di Informazione Finanziaria per l'Italia, *Rapporto Annuale 2023*, disponibile su: uif.bancaditalia.it.

[18] Sull'obbligo di iscrizione nella sezione speciale del registro tenuto dall'OAM, cfr. art. 17-bis, commi 8-bis e 8-ter, del D.Lgs. 13 agosto 2010, n. 141, introdotti dal D.Lgs. 25 maggio 2017, n. 90, e il D.M. del Ministero dell'economia e delle finanze 13 gennaio 2022.

sospettano o hanno motivi ragionevoli per sospettare che siano in corso o che siano state compiute o tentate operazioni di riciclaggio o di finanziamento del terrorismo»[19].

Dalla lettura sistematica della norma emergono tre elementi costitutivi fondamentali che ne definiscono il perimetro applicativo. Il primo è la soglia del sospetto, e non della certezza. Tant'è che il legislatore non richiede al soggetto obbligato, ovviamente, di condurre indagini di polizia giudiziaria né di possedere la prova certa della commissione di un reato. È infatti sufficiente il "dubbio fondato", ancorato a elementi di fatto oggettivi e a valutazioni logiche desumibili dalle informazioni acquisite in sede di adeguata verifica. La scelta di collocare la soglia al livello del sospetto ragionevole risponde a una precisa ratio deflattiva: se il soggetto obbligato avesse la certezza del reato, scatterebbe l'obbligo di denuncia all'Autorità giudiziaria; nell'incertezza fondata, la reazione doverosa è la SOS alla UIF. Il secondo elemento è il requisito della ragionevolezza: il sospetto non può essere frutto di mera intuizione soggettiva, ma deve essere corroborato da una valutazione logica e motivata dell'insieme delle informazioni disponibili sul cliente e sull'operazione. Il terzo elemento è l'estensione al tentativo: l'obbligo di segnalazione sussiste anche qualora l'operazione non sia stata eseguita o sia stata rifiutata dall'intermediario, poiché il tentativo stesso costituisce di per sé un forte indicatore di anomalia[20].

L'art. 35 specifica, inoltre, che il sospetto è desunto «dalle caratteristiche, dall'entità, dalla natura dell'operazione o da qualsiasi altra circostanza conosciuta in ragione delle funzioni esercitate, tenuto conto anche della capacità economica e dell'attività svolta dal soggetto cui è riferita». Si introduce qui il concetto di incoerenza come parametro valutativo centrale: un'operazione non è sospetta in termini assoluti, ma lo diviene in relazione al profilo economico-finanziario del cliente. Questa relativizzazione del sospetto impone al soggetto obbligato un'analisi contestuale e personalizzata, non riducibile all'applicazione meccanica di soglie quantitative predeterminate.

1.1.5 Il ruolo istituzionale della UIF e della Guardia di Finanza

Il sistema normativo ha disegnato un'architettura duale per la gestione dei flussi informativi generati dalle SOS, che vede la cooperazione tra due attori istituzionali con funzioni complementari e non sovrapponibili: l'Unità di Informazione Finanziaria per l'Italia (UIF) e il Nucleo Speciale di Polizia Valutaria (NSPV) della Guardia di Finanza.

L'UIF è istituita presso la Banca d'Italia dal D.Lgs. 231/2007, in sostituzione del previgente Ufficio Italiano dei Cambi; tale collocazione ne garantisce l'autonomia funzionale e l'indipendenza da ingerenze politiche. È essenziale precisare che l'UIF non è un organo di polizia giudiziaria: non dispone del potere di effettuare arresti, perquisizioni o intercettazioni, né può disporre sequestri, fatta eccezione per la sospensione amministrativa dell'operazione per un massimo di cinque giorni lavorativi ai sensi dell'art. 6, comma 4, lett. c), del decreto[21]. La funzione primaria della UIF è quella

[19] D.Lgs. 21 novembre 2007, n. 231, art. 35, comma 1.
[20] D.Lgs. 21 novembre 2007, n. 231, art. 35, comma 1 (riferimento al tentativo). Cfr. UIF, *Istruzioni per la rilevazione e la segnalazione delle operazioni sospette*, Provvedimento del 18 dicembre 2025, Parte Prima, Sezione IV, disponibili su: uif.bancaditalia.it.
[21] D.Lgs. 21 novembre 2007, n. 231, art. 6, comma 4, lett. c).

di agire da filtro qualificato tra il settore privato e l'autorità investigativa. Essa riceve le segnalazioni dai soggetti obbligati e le sottopone a un processo di analisi finanziaria attraverso strumenti proprietari, come il sistema RADAR (*Raccolta Analisi Dati Anti Riciclaggio*). Se rileva che determinate segnalazioni presentino indici di rischio significativi, le trasmetterà agli organi investigativi arricchite con un patrimonio informativo aggiuntivo.

Il destinatario principale di tale flusso è il Nucleo Speciale di Polizia Valutaria della Guardia di Finanza, che agisce come hub investigativo nazionale. Il NSPV non si limita a ricevere il rapporto della UIF, ma lo sottopone a un ulteriore vaglio operativo, incrociandolo con le banche dati di polizia e con l'anagrafe tributaria, sviluppando le segnalazioni avvalendosi dei poteri di polizia valutaria che consentono accessi bancari e ispezioni in via amministrativa, con una rapidità d'intervento essenziale per congelare situazioni a rischio[22]. Parallelamente, un flusso specifico è indirizzato alla Direzione Investigativa Antimafia (DIA), per le segnalazioni che presentino indici di rischio riconducibili a contesti di criminalità organizzata, dove la SOS assume una valenza strategica ulteriore: quella di strumento per l'attivazione delle misure di prevenzione patrimoniale[23].

È utile precisare, infine, che la SOS non costituisce prova penale in senso stretto, ma uno strumento di *intelligence* finanziaria con funzione endoprocedimentale: essa orienta le indagini, indica quindi "dove guardare", ma non può da sola fondare né un'accusa né una condanna. Gli organi investigativi sono tenuti a reperire riscontri probatori autonomi (estratti conto, fatture, intercettazioni) che costituiranno gli elementi oggettivi del fascicolo del Pubblico Ministero. In questo senso, la SOS funge da "innesco necessario ma non sufficiente" del procedimento investigativo, una soglia di attivazione che trasforma il sospetto soggettivo del banchiere in un impulso formale all'azione dello Stato.

Con il Pacchetto AML 2024 e la piena operatività dell'AMLA, il ruolo delle FIU nazionali, e quindi della UIF italiana, è destinato a evolversi ulteriormente: l'AMLA gestirà la rete FIU.net, potenziando lo scambio di informazioni tra le unità degli Stati membri e garantendo una visione transnazionale dei flussi finanziari sospetti che oggi rimane frammentata. Questo sviluppo istituzionale impatterà direttamente anche sulla qualità e sull'utilizzo delle SOS, aprendo nuove prospettive per l'analisi aggregata e cross-border delle anomalie finanziarie[24].

[22] I poteri di polizia valutaria del NSPV sono disciplinati dal D.P.R. 31 marzo 1988, n. 148 e dell'art. 9 del D.Lgs. 231/2007.
[23] D.Lgs. 21 novembre 2007, n. 231, art. 40, comma 1, lett. d) (trasmissione alla DIA e al NSPV della Guardia di Finanza).
[24] Regolamento (UE) 2024/1620, art. 47 (gestione di FIU.net e coordinamento delle Unità di Informazione Finanziaria nazionali da parte dell'AMLA).

1.2 Il processo attuale di individuazione dell'anomalia: monitoraggio automatico e valutazione discrezionale

1.2.1 La cornice normativa del processo di detection

Una volta delineato il quadro degli obblighi sostanziali, occorre analizzare il momento operativo in cui tali obblighi si traducono in attività concreta: il processo di individuazione (*detection*) delle anomalie. La normativa di vigilanza impone ai soggetti obbligati di dotarsi di presidi organizzativi e tecnologici adeguati a presidiare in modo integrato il rischio di riciclaggio. Il riferimento principale è costituito dalle Disposizioni di Banca d'Italia del 26 marzo 2019 in materia di organizzazione, procedure e controlli interni — più volte aggiornate dai provvedimenti del 1° agosto 2023, del 26 e 27 novembre 2024 e del 23 luglio 2025, sino alla versione consolidata diffusa nel 2026 — che declinano, sul piano organizzativo, i principi dell'approccio basato sul rischio[25].

Il sistema di rilevazione delle operazioni sospette si fonda su un'architettura duale, in cui convergono due canali paralleli e complementari: da un lato il monitoraggio automatico, affidato a procedure informatiche che analizzano massivamente i flussi finanziari; dall'altro la rilevazione umana, esercitata dalla rete commerciale a diretto contatto con la clientela. La compresenza di questi due canali non è casuale, ma riflette la natura intrinsecamente ibrida del giudizio di sospetto, che richiede tanto l'analisi quantitativa di grandi moli di dati quanto la valutazione qualitativa degli elementi comportamentali e relazionali. Come si avrà modo di argomentare, è proprio in questo rapporto tra automazione e discrezionalità umana che si colloca il filo conduttore della presente trattazione.

1.2.2 La procedura GIANOS e l'architettura del monitoraggio automatico

Nelle realtà bancarie di dimensioni significative, la mole di transazioni processate quotidianamente rende materialmente impossibile un controllo manuale esaustivo su ogni singola operazione. Ne consegue la necessità di ricorrere a software di monitoraggio che analizzano i flussi finanziari sulla base di regole predefinite, generando segnalazioni interne (c.d. *alert*) al verificarsi di determinate condizioni di anomalia.

Nel contesto italiano, lo strumento storicamente più diffuso è rappresentato dalla procedura GIANOS (*Generatore Indici di Anomalia per Operazioni Sospette*), sviluppata in ambito ABI (Associazione Bancaria Italiana) a partire dai primi anni Novanta, in risposta agli obblighi introdotti dal D.L. 3 maggio 1991, n. 143. È importante precisare la natura di tale strumento: GIANOS non è un mero software, bensì una metodologia complessa, finalizzata a oggettivare il rischio di riciclaggio attraverso l'elaborazione incrociata di dati statici e dinamici. È altresì rilevante sottolineare, sotto il profilo della tutela della riservatezza, che ciascun intermediario utilizza la procedura esclusivamente

[25] Banca d'Italia, *Disposizioni in materia di organizzazione, procedure e controlli interni volti a prevenire l'utilizzo degli intermediari a fini di riciclaggio e di finanziamento del terrorismo*, 26 marzo 2019, pubblicate in G.U. n. 83 dell'8 aprile 2019, e successive modificazioni (provvedimenti del 1° agosto 2023, 26 e 27 novembre 2024, 23 luglio 2025), disponibili su: bancaditalia.it.

sui dati della propria clientela: non esiste alcuna struttura centralizzata che aggreghi e consulti i dati di più banche.

Il funzionamento degli algoritmi di monitoraggio si fonda sulla costruzione di un profilo di rischio integrato per ciascun cliente, attraverso un processo di *risk assessment* automatico che sintetizza due macro-variabili fondamentali.

La prima è la componente statica, relativa al profilo soggettivo del cliente, elaborata al momento dell'instaurazione del rapporto continuativo (fase di *onboarding*). In questa fase l'algoritmo attribuisce un punteggio di rischio (*score*) sulla base di parametri anagrafici e giuridici predefiniti, quali la natura giuridica del soggetto (persona fisica, impresa individuale o società), il settore di attività economica e l'area geografica di residenza o di operatività prevalente. A titolo esemplificativo, un'impresa operante nel settore del "compro oro" o dell'edilizia, situata in un'area geografica ad alta densità criminale, riceverà automaticamente uno *score* di rischio intrinseco più elevato rispetto a un lavoratore dipendente residente in una zona a basso rischio, indipendentemente dall'operatività effettivamente posta in essere.

A questa componente statica si affianca la componente dinamica, che costituisce il vero cuore del monitoraggio algoritmico. Il motore inferenziale analizza i flussi finanziari del cliente confrontandoli con due parametri di riferimento: il comportamento storico del cliente medesimo e il comportamento medio del *cluster* di appartenenza. L'algoritmo ricerca, in termini statistici, lo scostamento (*outlier*) rispetto a un modello atteso. Qualora un cliente profilato come lavoratore dipendente con uno stipendio medio inizi improvvisamente a ricevere bonifici di importo decuplicato o a movimentare contante per cifre incompatibili con il reddito dichiarato, il sistema rileva un'incongruenza logica. Tale analisi si basa su soglie quantitative preimpostate, che tengono conto della frequenza delle operazioni, della loro direzione (entrata o uscita) e della tipologia di controparti coinvolte.

Particolare rilevanza assumono le regole di intercettazione degli schemi frazionati, tecnicamente definiti *smurfing*. Gli algoritmi sono programmati per aggregare operazioni che, singolarmente considerate, apparirebbero irrilevanti o al di sotto della soglia di legge — ad esempio molteplici versamenti di importo unitario contenuto effettuati in giorni consecutivi — ma che, valutate unitariamente, rivelano un intento elusivo. In tale ottica, GIANOS e i sistemi equivalenti operano cumulando i flussi su finestre temporali mobili (settimanali o mensili) e facendo scattare l'allarme non appena la somma algebrica delle transazioni superi i limiti di tolleranza parametrati al profilo del cliente.

1.2.3 Dall'alert alla valutazione: il problema dei falsi positivi

Il risultato di tale elaborazione è la generazione di un *alert*. È fondamentale sottolineare, tuttavia, che l'attivazione di un *alert* non coincide automaticamente con l'esistenza di un'operazione sospetta: essa segnala una semplice anomalia statistica, che necessita di un successivo vaglio qualitativo da parte di un analista. Si apre qui la problematica gestionale dei cosiddetti falsi positivi, ossia di quelle evidenze generate dall'algoritmo che, a una successiva analisi umana, si rivelano prive di fondamento illecito — si pensi, ad esempio, a un picco di operatività riconducibile alla vendita di un immobile regolarmente rogitato.

L'efficienza di un sistema di monitoraggio si misura proprio nella capacità di tarare gli algoritmi in modo da minimizzare il "rumore" di fondo dei falsi positivi, consentendo agli analisti della Funzione Antiriciclaggio di concentrare le risorse investigative sulle anomalie che presentano un reale sostrato di rischio. Le stime disponibili indicano che nei sistemi tradizionali basati su regole rigide oltre il novanta per cento degli *alert* generati si rivela infondato[26], con un conseguente dispendio di risorse che rappresenta una delle principali inefficienze dell'attuale assetto. Tale criticità costituisce, peraltro, una delle ragioni principali che spingono il settore verso l'adozione di tecniche di *machine learning*, capaci di affinare progressivamente la capacità discriminante del sistema.

1.2.4 L'evoluzione verso logiche predittive

Negli scenari più evoluti, gli intermediari hanno iniziato a integrare le tradizionali logiche deterministiche — fondate su regole del tipo "se… allora" — con approcci basati sull'intelligenza artificiale e sul *machine learning*. A differenza dei sistemi tradizionali, che applicano regole statiche, i modelli di apprendimento automatico sono in grado di imparare dai dati storici delle segnalazioni passate, identificando correlazioni non lineari e *pattern* comportamentali complessi che sfuggirebbero a una programmazione classica. Ciò consente di far evolvere il monitoraggio da una dimensione puramente reattiva a una parzialmente predittiva, anticipando potenziali scenari di rischio sulla base di deboli segnali premonitori presenti nell'operatività della clientela.

Si tratta di un'evoluzione di portata strutturale, che sarà oggetto di analisi approfondita nel Capitolo 2. Intanto è sufficiente anticipare come il passaggio dal paradigma *rule-based* a quello predittivo non elimini, bensì ridefinisca, il ruolo del fattore umano: l'algoritmo, per quanto sofisticato, continua a generare evidenze che necessitano di una validazione qualitativa, e la decisione finale in merito alla segnalazione rimane irriducibilmente affidata al giudizio del delegato SOS. La tecnologia ottimizza il processo di rilevazione, ma non ne assume la responsabilità giuridica.

1.2.5 La rilevazione umana: il presidio insostituibile della rete

Accanto al canale automatizzato, il presidio umano esercitato dalla rete commerciale (filiali, consulenti, *private banker*) rimane una componente insostituibile del processo di individuazione. Se l'algoritmo eccelle nell'analisi quantitativa e nella gestione di grandi moli di dati strutturati, esso risulta inefficace di fronte agli aspetti qualitativi e relazionali che spesso costituiscono i primi segnali di un'attività illecita. L'operatore di sportello e il gestore della relazione non agiscono come meri esecutori di transazioni, ma rappresentano la prima linea di difesa del sistema, chiamati a esercitare quella sensibilità critica che la dottrina definisce scetticismo professionale[27].

---

[26] LANNOO K. – PARLOUR R., *Anti-Money Laundering in the EU: Time to get serious*, CEPS-ECRI Task Force Report, Bruxelles, gennaio 2021, disponibile su: ceps.eu, dove si osserva che i falsi positivi rappresentano spesso il 95-98 per cento degli alert generati dai sistemi di monitoraggio.

[27] Il concetto di *scetticismo professionale* è mutuato dalla revisione contabile: cfr. *Principio di revisione internazionale* (ISA Italia) 200, Obiettivi generali del revisore indipendente e svolgimento della revisione contabile in conformità ai principi di revisione internazionali, par. 13, lett. l) e par. 15.

Il valore aggiunto della rilevazione umana risiede nella capacità di captare le cosiddette *soft information*, ossia quell'insieme di dati non codificabili che emergono durante l'interazione diretta con la clientela. Un indicatore primario di anomalia è riscontrabile nell'atteggiamento comportamentale del cliente al momento dell'operazione: l'operatore è tenuto a osservare se il soggetto manifesti un ingiustificato nervosismo, una fretta eccessiva nel concludere la transazione o una palese reticenza nel fornire le informazioni necessarie all'adeguata verifica. Risposte evasive, generiche o palesemente preconfezionate di fronte a domande standard sulla provenienza dei fondi costituiscono un campanello d'allarme che nessun software potrebbe intercettare con la medesima efficacia[28].

Particolare attenzione deve essere posta alle dinamiche relazionali che suggeriscono l'interposizione fittizia di persona, fenomeno noto come utilizzo di "prestanome" o "teste di legno". L'esperienza operativa dimostra che il riciclatore professionista tende a non esporsi in prima persona, servendosi di soggetti terzi spesso privi di adeguata cultura finanziaria o in stato di bisogno. In tale contesto, l'operatore deve saper cogliere le incongruenze tra il profilo apparente del cliente e la sua operatività reale: una casistica frequente riguarda il cliente che si presenti accompagnato da un "consulente" non titolato che, di fatto, dirige l'operazione, suggerisce le risposte o detiene fisicamente la documentazione e i dispositivi di accesso ai conti. Tale situazione costituisce un segnale inequivocabile che il soggetto allo sportello potrebbe non essere il titolare effettivo reale dei fondi, bensì un mero strumento di schermatura dell'intento criminale[29].

Il presidio umano è altresì fondamentale per valutare la coerenza logica della documentazione giustificativa. Mentre un sistema informatico può limitarsi a registrare la presenza di una fattura o di un contratto, l'occhio umano è in grado di valutarne la genuinità sostanziale: anomalie quali fatture con descrizioni di servizi generiche e incongruenti rispetto all'oggetto sociale dell'azienda, contratti privi di data certa o documenti contabili palesemente alterati emergono solo attraverso un esame critico manuale[30].

È infine doveroso evidenziare la tensione strutturale cui è sottoposta la rete commerciale, stretta tra obiettivi di *budget* e obblighi di *compliance*. Il successo del sistema di segnalazione dipende largamente dalla cultura aziendale: laddove l'intermediario incentivi esclusivamente la vendita, l'operatore tenderà a sottovalutare i segnali di rischio per non compromettere la relazione commerciale. Un efficace sistema di controlli interni deve, al contrario, formare il personale affinché comprenda che la segnalazione di un'operazione sospetta non costituisce un atto di delazione, bensì un obbligo giuridico posto a tutela dell'integrità stessa dell'intermediario. La capacità di porre la "domanda giusta" al momento giusto, osservando la reazione non verbale del cliente, rimane dunque una competenza distintiva. Questa abilità dell'operatore separa una *compliance*

[28] Sulle *soft information*, contrapposte alle informazioni codificabili in forma numerica (*hard infomation*), cfr. LIBERTI J.M. – PETERSEN M.A., «Information: Hard and Soft», in *The Review of Corporate Finance Studies*, vol. 8, n.1, 2019.
[29] Sull'interposizione fittizia e sull'individuazione del titolare effettivo reale, cfr. UIF, *Indicatori di anomalia*, Provvedimento del 12 maggio 2023, disponibile su: uif.bancaditalia.it
[30] *Ibidem*.

sostanziale da una meramente formale; competenza che, come si argomenterà, nessuna automazione è in grado di replicare integralmente.

1.3 Le criticità odierne: il fenomeno del Defensive Reporting e l'elevato numero di falsi positivi.

1.3.1 Il sovraccarico informativo come patologia del sistema

L'analisi condotta nei paragrafi precedenti ha delineato un sistema di prevenzione dotato di un'architettura normativa sofisticata e di presidi tecnologici e organizzativi articolati. Tuttavia, la trasposizione operativa di tali obblighi nelle dinamiche quotidiane degli intermediari ha fatto emergere, nel corso degli anni, alcune criticità strutturali che rischiano di comprometterne l'efficacia. Il dato quantitativo offre una prima misura del fenomeno: nel 2024 la UIF ha ricevuto 145.401 segnalazioni di operazioni sospette, a fronte delle 150.418 del 2023 e di un volume che, nel decennio precedente, era cresciuto in modo pressoché ininterrotto[31]. Dopo la lieve contrazione del biennio 2023-2024 le segnalazioni sono tornate a crescere nel 2025, e il sistema continua a produrre un flusso segnaletico di dimensioni imponenti, che pone interrogativi sulla sua sostenibilità e sull'effettiva qualità dell'informazione trasmessa.

Si assiste, quindi, a una tensione strutturale tra l'esigenza di esaustività — ossia la necessità di intercettare ogni possibile anomalia — e il rischio di un sovraccarico informativo (*information overload*) che, paradossalmente, può depotenziare l'efficacia complessiva dell'azione investigativa. Una mole segnaletica eccessiva, infatti, disperde le risorse analitiche delle autorità in un mare di comunicazioni a basso valore informativo, rendendo più arduo individuare le operazioni realmente meritevoli di approfondimento. Il paradosso è evidente: l'accumulo indiscriminato di segnalazioni, lungi dal rafforzare la capacità di contrasto, rischia di comprometterla, secondo una logica per cui "segnalare tutto equivale a non segnalare nulla".

1.3.2 Il *Defensive Reporting*: genesi e razionalità economica

La causa principale di tale sovraccarico è comunemente identificata nel fenomeno del *defensive reporting* (segnalazione difensiva). Con tale espressione si designa la tendenza dei soggetti obbligati — e in particolare degli intermediari bancari — a trasmettere segnalazioni alla UIF anche a fronte di sospetti deboli, generici o scarsamente fondati, al solo scopo di tutelarsi dal rischio sanzionatorio.

La *ratio* di tale comportamento non è irrazionale, bensì risponde a una precisa logica di analisi costi-benefici. Per l'intermediario, il costo marginale connesso all'invio di una segnalazione superflua è prossimo allo zero. Per contro, il costo potenziale derivante dalla mancata segnalazione di un'operazione che si riveli successivamente collegata a un fatto di riciclaggio può essere assai elevato: le sanzioni amministrative pecuniarie previste dal

[31] I dati sono tratti da UIF, *Rapporto annuale per il 2024, n. 17 – 2025*, disponibile su: uif.bancaditalia.it. Per il dato 2023 (150.418 SOS) cfr. UIF, *Rapporto annuale per il 2023, n. 16 – 2024*, ivi.

D.Lgs. 231/2007 (art. 58 e, per gli intermediari vigilati, art. 62) raggiungono importi di rilievo, cui si aggiungono le conseguenze sul piano reputazionale e i possibili riflessi sulla posizione degli esponenti aziendali[32]. Di fronte a tale asimmetria, l'intermediario, stretto tra l'incudine della vigilanza ispettiva e il martello del rischio sanzionatorio, opta razionalmente per un approccio di massima precauzione, inoltrando all'Autorità ogni evidenza dubbia.

Il risultato aggregato di questo comportamento individualmente razionale è, tuttavia, collettivamente subottimale: la UIF si trova a dover gestire una massa di segnalazioni a bassa qualità informativa, che diluisce le risorse analitiche e ostacola l'individuazione delle operazioni realmente rilevanti. Si configura, in tal senso, una tipica situazione di esternalità negativa, in cui la somma di scelte prudenziali individuali produce un deterioramento della qualità complessiva del sistema.

Il fenomeno è stato oggetto di un'attenzione specifica da parte della letteratura economica, che ne ha fornito una formalizzazione rigorosa. Il contributo di riferimento è quello di Takáts, che per primo ha analizzato l'enforcement antiriciclaggio come un problema di agenzia fra le banche, che dispongono delle informazioni sulle operazioni, e le autorità, che di quelle informazioni hanno bisogno per indirizzare le indagini[33]. Il modello mostra che sanzioni eccessive per l'omessa segnalazione inducono le banche a segnalare anche operazioni poco sospette, e che questo eccesso finisce per diluire il valore informativo dell'insieme delle segnalazioni: è il fenomeno che l'autore definisce, con un'immagine efficace, del gridare «al lupo». Le misure correttive suggerite vanno in una direzione controintuitiva rispetto alla logica puramente repressiva, e cioè la riduzione delle sanzioni e l'introduzione di un costo per la segnalazione, così da riequilibrare gli incentivi.

Sulla stessa linea si colloca l'analisi comparata di Unger e van Waarden, che mette a confronto un sistema fondato su regole e uno basato sul rischio (gli Stati Uniti e i Paesi Bassi) proprio sotto il profilo del rischio di affogare nei dati: al forte aumento delle segnalazioni statunitensi corrisponde un calo della loro qualità, mentre il sistema olandese, che ne produce molte meno, registra un contenuto informativo più elevato.[34].

Il caso italiano è stato indagato empiricamente in due lavori. Il primo, condotto nell'ambito della stessa UIF, ha stimato un modello econometrico delle segnalazioni inviate da ciascuna banca nelle singole province, in funzione dell'operatività, del rischio di riciclaggio e dell'attività economica locale, con l'obiettivo di individuare gli intermediari che segnalano meno (o più) di quanto ci si attenderebbe, e di orientare di conseguenza i controlli[35]. Il secondo ha mostrato, sul piano teorico, che può emergere un *trade-off* fra accuratezza delle segnalazioni, intesa come minimizzazione degli errori di

[32] D.Lgs. 21 novembre 2007, n. 231, art. 58 e art 62.
[33] TAKÁTS E., «A Theory of "Crying Wolf": The Economics of Money Laundering Enforcement», in *The Journal of Law, Economics, and Organization*, vol. 27, n. 1, 2011.
[34] UNGER B. – VAN WAARDEN F., «How to Dodge Drowning in Data? Rule- and Risk-Based Anti Money Laundering Policies Compared», in *Review of Law & Economics*, vol. 5, n. 2, 2009.
[35] GARA M. – PAUSELLI C., «Looking at 'Crying Wolf' from a Different Perspective: An Attempt at Detecting Banks Under- and Over-Reporting of Suspicious Transactions», in *Italian Economic Journal*, vol. 6, n. 2, 2020. Lo studio è stato pubblicato originariamente nella collana UIF *Quaderni dell'antiriciclaggio – Analisi e studi*, n. 4.

primo e di secondo tipo, e deterrenza; e, sul piano empirico, che dopo l'introduzione dell'approccio basato sul rischio in Italia il sistema tende a massimizzare la deterrenza a scapito dell'accuratezza, con dati che suggeriscono come i banchieri italiani siano propensi a gridare «al lupo»[36].

1.3.3 L'inefficienza economica e il paradosso dei falsi positivi

Strettamente connesso al *defensive reporting*, ma logicamente distinto, è il problema dei falsi positivi. Mentre la segnalazione difensiva attiene alla scelta consapevole dell'intermediario di segnalare "per cautela", il falso positivo è un fenomeno che origina a monte, nel funzionamento stesso dei sistemi di monitoraggio automatico esaminati nel paragrafo 1.2.

Come si è visto, i sistemi tradizionali operano prevalentemente secondo logiche deterministiche (*rule-based*): al superamento di una determinata soglia o al verificarsi di una certa condizione parametrica, il sistema genera un *alert*. La rigidità di tali regole comporta, tuttavia, che la stragrande maggioranza degli allarmi generati non corrisponda a reali tentativi di riciclaggio, bensì a operatività lecita che, per mera coincidenza statistica, attiva i criteri di anomalia.

Le implicazioni economiche di tale inefficienza sono considerevoli. L'apparato di prevenzione rappresenta per gli intermediari una voce di costo imponente e in costante crescita, tecnicamente definita *cost of compliance*, che non si esaurisce negli investimenti in infrastrutture informatiche, ma assorbe prevalentemente risorse umane ad alta specializzazione dedicate all'analisi manuale delle posizioni anomale. Il fatto che tali risorse siano impiegate, per la quasi totalità del loro tempo, nell'eliminazione di allarmi infondati configura un'allocazione inefficiente, che sottrae tempo prezioso all'investigazione dei pochi casi effettivamente complessi e rischiosi. Ne deriva un elevato costo-opportunità per il sistema finanziario e, in ultima analisi, per la collettività, giacché i costi di *compliance* vengono inevitabilmente traslati sulla clientela sotto forma di maggiori commissioni, senza che a ciò corrisponda un proporzionale incremento della sicurezza effettiva del sistema.

1.3.4 La carenza di *feedback* e l'impossibilità di taratura

Una terza criticità, ripetutamente denunciata dalle associazioni rappresentative dei soggetti obbligati, e in particolare dall'ABI, attiene alla carenza di un adeguato ritorno informativo (*feedback*) da parte delle autorità sugli esiti delle segnalazioni trasmesse. Storicamente, il flusso informativo si è configurato come prevalentemente unidirezionale: gli intermediari trasmettono i dati alla UIF, ma solo di rado ricevono indicazioni sull'utilità di quanto segnalato. L'intermediario, pertanto, tende a ignorare se una determinata segnalazione abbia contribuito all'individuazione di un fenomeno criminale,

[36] DALLA PELLEGRINA L. – DI MAIO G. – MASCIANDARO D. – SARACENO M., «Are Bankers "Crying Wolf"? Type I, Type II Errors and Deterrence in Anti-Money Laundering: The Italian Case», in *Italian Economic Journal*, vol. 9, n. 2, 2023.

sia stata archiviata in quanto infondata, o sia stata classificata come irrilevante per scarsa qualità.

Tale assenza di ritorno informativo genera un duplice ordine di problemi. Sul piano tecnico, essa impedisce la taratura dei sistemi: senza conoscere quali segnalazioni si siano rivelate utili e quali no, la Funzione Antiriciclaggio non è in grado di affinare i propri algoritmi di monitoraggio né di calibrare la formazione del personale, con il rischio di perpetuare all'infinito i medesimi errori. Sul piano motivazionale, il personale della rete commerciale finisce per percepire l'adempimento segnaletico come un mero onere burocratico, privo di riscontro e di valore percepito, con conseguente disaffezione verso la funzione di presidio.

Occorre tuttavia segnalare che, proprio su questo fronte, si registrano sviluppi significativi. La UIF ha avviato un progetto strutturato per il monitoraggio della qualità della collaborazione attiva e ha progressivamente potenziato i flussi di ritorno verso i segnalanti. Come si avrà modo di approfondire nel paragrafo 1.4, le nuove Istruzioni UIF del 18 dicembre 2025 dedicano la Sezione VIII della Parte Prima al *flusso di ritorno*, prevedendo comunicazioni almeno semestrali sugli esiti delle segnalazioni e schede di *feedback* a cadenza annuale: un intervento che risponde direttamente alla criticità qui descritta e che testimonia la progressiva evoluzione del sistema verso una logica di collaborazione bidirezionale[37].

1.3.5 L'andamento recente delle segnalazioni

L'andamento delle segnalazioni negli ultimi anni non segue una direzione lineare (Figura 1). Dopo due esercizi di flessione, con una contrazione del 3,2 per cento nel 2023 e del 3,3 per cento nel 2024, anno chiuso a 145.401 segnalazioni[38], il 2025 ha registrato una netta ripresa: 162.059 segnalazioni ricevute, pari a un incremento dell'11,5 per cento, con un valore complessivo delle operazioni segnalate superiore ai novantacinque miliardi di euro[39]. La crescita è proseguita nei primi mesi dell'anno successivo.

Le ragioni dell'aumento, secondo quanto riferisce la stessa Unità, sono in larga parte riconducibili a fattori esterni al comportamento dei segnalanti storici. Pesano l'ingresso nel perimetro di operatori di recente iscrizione, in particolare banche telematiche specializzate nell'operatività online, e soprattutto l'espansione delle segnalazioni collegate a truffe e frodi informatiche, che da sole si attestano intorno alle trentunomilaseicento unità. Significativa anche la crescita degli operatori non finanziari, fra i quali i prestatori di servizi in valuta virtuale hanno quasi raddoppiato il proprio contributo[40]. Il dato riflette dunque un mutamento nella fisionomia dei fenomeni criminali e nella composizione della platea dei soggetti obbligati, più che una variazione nella propensione a segnalare.

---

[37] UIF, *Istruzioni per la rilevazione e la segnalazione delle operazioni sospette*, Provvedimento del 18 dicembre 2025, Parte Prima, Sezione VIII (Flusso di ritorno della UIF), in vigore dal 1° luglio 2026, disponibile su: uif.bancaditalia.it.
[38] UIF, *Rapporto annuale per il 2024*, n. 17, 2025, disponibile su: uif.bancaditalia.it.
[39] UIF, *Rapporto annuale per il 2025*, n. 18, 2026, disponibile su: uif.bancaditalia.it.
[40] UIF, *Rapporto annuale per il 2025*, cit.

Sul versante qualitativo il quadro è composito. La quota di segnalazioni classificate a rischio basso o nullo prosegue la propria discesa, collocandosi al 18,9 per cento a fronte del 20,4 dell'anno precedente, e migliora la tempestività dell'invio. Al tempo stesso, però, si riduce la quota delle segnalazioni che l'Unità colloca nelle fasce di rischio medio-alto e alto, passata dal 45,9 al 42,1 per cento, mentre si espande in misura consistente la fascia intermedia, dal 32,9 al 38,8 per cento[41]. Anche in questo caso la spiegazione rimanda ai fenomeni emergenti: le segnalazioni connesse a truffe informatiche, criptoattività e operatori di pagamento dai profili atipici tendono a collocarsi in una zona di rischio intermedia.

Emerge da qui un'indicazione di metodo. Il volume delle segnalazioni non costituisce, di per sé, un indicatore dell'efficacia del sistema. Una crescita può riflettere l'emergere di nuove tipologie criminali o l'ampliamento della platea dei segnalanti; una flessione può derivare da fattori congiunturali. Volume e qualità sono grandezze distinte, che rispondono a logiche proprie e che vanno lette separatamente.

La distinzione ha una conseguenza che interessa il seguito della trattazione. Se il *defensive reporting* nasce, come si è argomentato, da un'asimmetria negli incentivi (costo trascurabile della segnalazione superflua e costo elevato dell'omissione), esso è un fenomeno strutturale, che l'andamento dei volumi non misura e che il solo progresso tecnologico non corregge. Un sistema di monitoraggio più preciso riduce il numero di allarmi infondati che raggiungono l'analista, ma lascia intatto il calcolo che induce l'intermediario a segnalare nel dubbio. È un punto che accompagnerà l'esame delle soluzioni tecnologiche nel capitolo seguente, e che invita a distinguere ciò che la tecnologia può risolvere da ciò che dipende dall'architettura degli incentivi.

Le criticità sin qui esaminate, va aggiunto, convivono con una valutazione complessivamente positiva del sistema italiano. All'inizio del 2026 si è conclusa la valutazione mutuale condotta dal GAFI, che ha attestato la sostanziale efficacia dell'apparato nazionale di prevenzione e contrasto, riconoscendo in particolare il ruolo della UIF nella produzione di analisi finanziaria di qualità elevata[42]. Le inefficienze descritte non ne compromettono la tenuta: ne indicano i margini di miglioramento, ed è su quei margini che si innesta la riflessione sulle nuove tecnologie.

[41] UIF, *Rapporto annuale per il 2025*, cit. La classificazione per fasce di rischio è operata dall'Unità in esito al processo di analisi delle segnalazioni ricevute.
[42] UIF, *Rapporto annuale per il 2025*, cit.

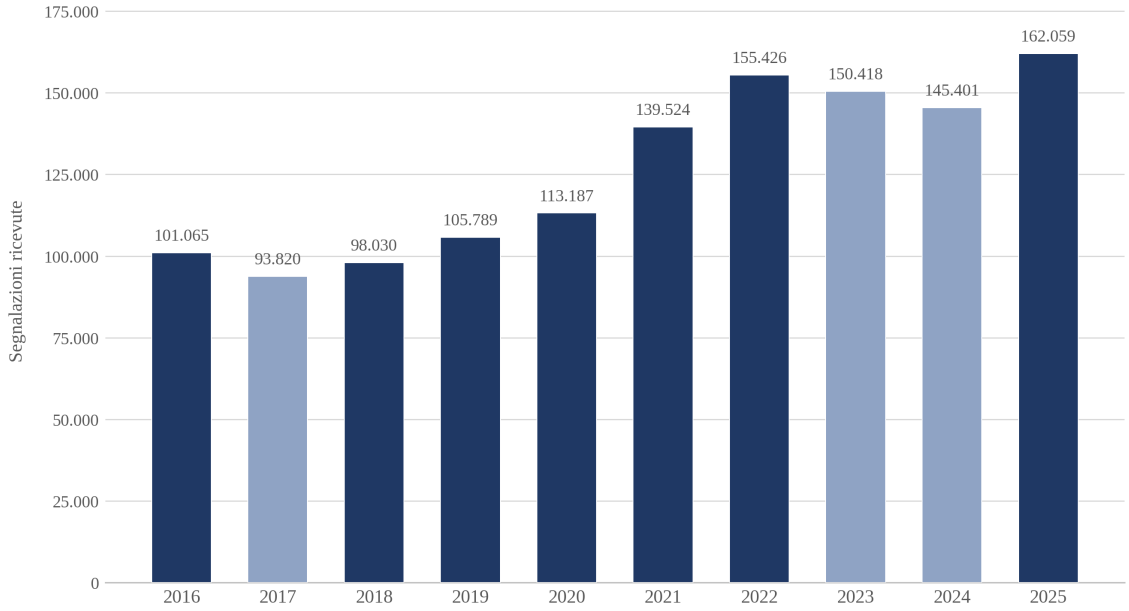

Figura 1 – Segnalazioni di operazioni sospette ricevute dalla UIF (2016-2025). In tonalità più chiara gli anni di flessione. Fonte: elaborazione propria su dati UIF (Quaderni dell'antiriciclaggio, Dati statistici e Rapporti annuali).

1.4 Le nuove Istruzioni UIF del 18 dicembre 2025: la disciplina rinnovata della collaborazione attiva

1.4.1 Contesto e ratio del nuovo Provvedimento

La disamina delle criticità operative condotta nel paragrafo precedente trova un riscontro diretto e attualissimo nel più recente intervento regolatorio della UIF in materia. Con Provvedimento del 18 dicembre 2025, l'Unità di Informazione Finanziaria per l'Italia ha emanato le nuove Istruzioni per la rilevazione e la segnalazione delle operazioni sospette, destinate a entrare in vigore il 1° luglio 2026 (data che coincide con il periodo di redazione del presente elaborato) in sostituzione del precedente Provvedimento del 4 maggio 2011 e dei relativi allegati[43]. Si tratta, pertanto, di una fonte di primissima attualità, la cui analisi consente di verificare come il legislatore secondario stia rispondendo, in concreto, alle disfunzioni sistemiche sopra descritte.

Il nuovo Provvedimento trova il proprio fondamento normativo negli articoli 6, comma 4, lettera d), e 35, comma 3, del D.Lgs. 231/2007, che attribuiscono alla UIF il compito di emanare istruzioni sui dati e sulle informazioni da inserire nelle segnalazioni, sulla relativa tempistica e sulle modalità di tutela della riservatezza del segnalante[44]. Sul piano metodologico, è significativo rilevare come i contenuti del documento siano stati elaborati

[43] UIF – Unità di Informazione Finanziaria per l'Italia, Istruzioni per la rilevazione e la segnalazione delle operazioni sospette, Provvedimento del 18 dicembre 2025, in vigore dal 1° luglio 2026, Disposizioni finali, disponibile su: uif.bancaditalia.it. Il medesimo Provvedimento dispone che, a decorrere dalla stessa data, cessa di applicarsi il Provvedimento della UIF del 4 maggio 2011.
[44] D.Lgs. 21 novembre 2007, n. 231, art. 6, comma 4, lett. d) e art. 35, comma 3.

in collaborazione con la Guardia di Finanza, tenuto conto delle interlocuzioni con la Direzione Investigativa Antimafia, le Autorità di vigilanza di settore e gli Organismi di autoregolamentazione, nonché delle osservazioni raccolte in sede di consultazione pubblica: un metodo partecipativo che riflette la natura sistemica e collaborativa della funzione antiriciclaggio[45].

Un ulteriore elemento di rilievo è il fatto che le nuove Istruzioni richiamino espressamente, tra le proprie fonti di riferimento, il Pacchetto AML 2024 — ossia il Regolamento (UE) 2024/1624, il Regolamento (UE) 2024/1620 istitutivo dell'AMLA e la Direttiva (UE) 2024/1640 — nonché, aspetto di particolare interesse per la presente trattazione, il Regolamento (UE) 2024/1689 (AI Act)[46]. Il documento si pone dunque consapevolmente all'incrocio tra la riforma dell'architettura antiriciclaggio europea e la nascente regolamentazione dell'intelligenza artificiale, confermando l'attualità del tema al centro di questa tesi.

1.4.2 Il rifiuto degli automatismi come risposta al *Defensive Reporting*

Il profilo di maggiore interesse del nuovo Provvedimento, in relazione alle criticità esaminate in precedenza, risiede nella netta presa di posizione contro gli automatismi segnaletici. Sin dai principi generali, le Istruzioni stabiliscono che, nel processo valutativo, devono essere evitati automatismi e approcci cautelativi. Per esempio, processi fondati sul mero superamento di soglie quantitative minime, sulla ricezione di richieste di informazioni o sui controlli delle autorità[47]. La formulazione è inequivocabile e costituisce una risposta diretta al fenomeno del *defensive reporting*: la UIF non si limita a scoraggiare la segnalazione difensiva, ma ne individua puntualmente le cause tipiche, escludendo che esse possano, di per sé, fondare l'obbligo di segnalazione.

La Sezione IV della Parte Prima specifica ulteriormente questo principio, elencando una serie di circostanze che, singolarmente considerate, non sono sufficienti a giustificare una segnalazione: tra queste, il rischio elevato attribuito al soggetto, l'individuazione di notizie negative sul medesimo, la ricezione di richieste informative da parte delle autorità competenti e la mera ricorrenza di misure penali a suo carico[48]. Analogamente, il ricorso a operazioni in contanti, anche se reiterato e a prescindere dal superamento delle soglie di legge, non integra di per sé un presupposto di segnalazione[49]. Il messaggio che ne emerge è coerente e sistematico: la segnalazione deve essere l'esito di un processo valutativo ragionato e contestualizzato, non il prodotto meccanico del superamento di una soglia o dell'attivazione di un indicatore.

Tale impostazione contribuisce a spiegare, sul piano interpretativo, la contrazione qualitativamente selettiva del flusso segnaletico registrata nel biennio 2023-2024 e

[45] UIF, Istruzioni del 18 dicembre 2025, cit., Premessa.
[46] Ivi, Disposizioni preliminari, Sezione I (Fonti normative), che richiama tra l'altro il Regolamento (UE) 2024/1620, il Regolamento (UE) 2024/1624, la Direttiva (UE) 2024/1640 e il Regolamento (UE) 2024/1689.
[47] Ivi, Parte Prima, Sezione I (Principi generali).
[48] Ivi, Parte Prima, Sezione IV (Segnalazione delle operazioni sospette).
[49] Ivi, Parte Prima, Sezione IV; cfr. altresì l'art. 4, comma 5, lett. d), del Provvedimento della UIF del 12 maggio 2023.

commentata nel paragrafo precedente: la progressiva riduzione delle segnalazioni a basso rischio riflette proprio quel superamento della logica difensiva che il nuovo Provvedimento intende consolidare sul piano normativo[50].

1.4.3 Il processo valutativo in tre fasi e la tracciabilità della decisione

Le nuove Istruzioni formalizzano il processo di collaborazione attiva articolandolo in tre fasi logicamente distinte: l'individuazione delle anomalie, il loro esame e la segnalazione vera e propria[51]. Con riferimento alla fase di esame, il Provvedimento consente ai destinatari di dimensioni maggiori di prevedere un vaglio preliminare, condotto da un soggetto o da una struttura dedicata, volto a escludere le anomalie che, a una prima analisi, risultino agevolmente giustificabili; solo le anomalie che persistono dopo tale filtro sono sottoposte alla valutazione conclusiva[52].

Particolarmente rilevante, anche in prospettiva del successivo discorso sull'intelligenza artificiale, è la disciplina della decisione di esclusione del sospetto. Le Istruzioni chiariscono che il processo valutativo può legittimamente concludersi con l'archiviazione interna, anche in presenza di fattispecie astrattamente riconducibili a indicatori di anomalia, qualora l'analisi complessiva induca a scartare il sospetto[53]. In tale evenienza, tuttavia, il destinatario è tenuto ad adottare accorgimenti idonei ad agevolare la ricostruzione a posteriori delle valutazioni effettuate, conservando una traccia, anche sintetica, delle ragioni ritenute sufficienti per escludere il sospetto[54].

1.4.4 L'apertura all'intelligenza artificiale e la centralità dell'intervento umano

Il passaggio di maggiore significato per l'impianto argomentativo della presente tesi è contenuto nella Sezione II della Parte Prima, dedicata all'individuazione delle anomalie. In tale sede, per la prima volta in un provvedimento della UIF, si riconosce espressamente che, tra gli strumenti informatici di selezione delle operatività anomale, rientrano anche quelli basati su sistemi di intelligenza artificiale[55].

L'apertura non è, tuttavia, incondizionata. Il Provvedimento subordina l'impiego di tali sistemi a tre precise condizioni: essi devono essere conformi alle disposizioni a essi eventualmente applicabili — con un rinvio esplicito, in nota, al Regolamento (UE) 2024/1689 (AI Act) —, devono basarsi su dati oggettivi e verificabili ed essere accompagnati da adeguate valutazioni svolte con l'intervento umano, al fine di controllare ed eventualmente validare le anomalie da essi evidenziate[56]. Si tratta di una formulazione di straordinaria rilevanza per il presente lavoro, in quanto cristallizza, sul piano della fonte

[50] Il dato sulla distribuzione delle SOS per classe di rischio è tratto da UIF, Rapporto annuale per il 2024, n. 17 – 2025.
[51] UIF, Istruzioni del 18 dicembre 2025, cit., Parte Prima, Sezioni II, III e IV.
[52] Ivi, Parte Prima, Sezione III (Esame delle anomalie).
[53] Ivi, Parte Prima, Sezione III.
[54] Ivi, Parte Prima, Sezione III.
[55] Ivi, Parte Prima, Sezione II (Individuazione delle anomalie).
[56] Ivi, Parte Prima, Sezione II, ove il rinvio all'AI Act è contenuto in nota al testo del Provvedimento. Il riferimento è al Regolamento (UE) 2024/1689.

secondaria, il principio che ne costituisce il filo conduttore: l'intelligenza artificiale è ammessa quale strumento di supporto all'individuazione delle anomalie, ma la sua produzione non può tradursi automaticamente in una segnalazione, dovendo essere sempre sottoposta al controllo e alla validazione del decisore umano.

1.4.5 Il rafforzamento del flusso di ritorno

Un ultimo profilo meritevole di attenzione, in diretta continuità con la criticità del *feedback* esaminata nel paragrafo 1.3, è costituito dalla Sezione VIII della Parte Prima, interamente dedicata al flusso di ritorno della UIF. Il Provvedimento istituzionalizza un sistema strutturato di comunicazioni verso i destinatari, articolato su più livelli[57].

In primo luogo, la UIF si impegna a comunicare gli esiti delle segnalazioni con cadenza almeno semestrale, distinguendo tra le segnalazioni per le quali non ha ravvisato sufficienti elementi di rischio a supporto del sospetto (il cosiddetto elenco A) e quelle che, pur presentando alcuni elementi a sostegno, sono state classificate a basso rischio (elenco B)[58]. In secondo luogo, è prevista una scheda di *feedback* a cadenza almeno annuale, contenente valutazioni di sintesi sulla qualità della collaborazione attiva del destinatario, con riferimento al grado di partecipazione, alla tempestività, alla completezza della compilazione e alla qualità complessiva delle segnalazioni[59].

L'introduzione di un simile meccanismo segna il passaggio, tanto auspicato dalle associazioni di categoria, da un flusso informativo unidirezionale a una logica di collaborazione bidirezionale e circolare, in cui il dato non viaggia più soltanto dall'intermediario verso l'Autorità, ma ritorna al segnalante sotto forma di indicazione qualitativa. Tale ritorno informativo consente agli intermediari di affinare progressivamente i propri processi di individuazione, realizzando quel circolo virtuoso di miglioramento continuo che rappresenta la condizione necessaria per innalzare la qualità complessiva del sistema.

1.4.6 Conclusioni del quadro normativo

L'analisi delle nuove Istruzioni UIF consente di tracciare un bilancio del quadro delineato nel presente capitolo. Il sistema delle segnalazioni di operazioni sospette si presenta oggi come un meccanismo maturo sul piano normativo e istituzionale, ma ancora gravato da inefficienze strutturali — il sovraccarico informativo, il *defensive reporting*, l'elevato tasso di falsi positivi, la storica carenza di *feedback* — che ne limitano l'efficacia. Il Provvedimento del 18 dicembre 2025 rappresenta la più recente e organica risposta del regolatore a tali criticità, muovendosi lungo tre direttrici convergenti: la lotta agli automatismi segnaletici, l'apertura controllata all'intelligenza artificiale e il rafforzamento del ritorno informativo.

È proprio nella seconda di queste direttrici — l'ammissione dell'intelligenza artificiale quale strumento di supporto, subordinata al controllo umano — che si colloca il nucleo

[57] UIF, Istruzioni del 18 dicembre 2025, cit., Parte Prima, Sezione VIII (Flusso di ritorno della UIF).
[58] Ivi, Parte Prima, Sezione VIII.
[59] Ivi, Parte Prima, Sezione VIII.

problematico del capitolo successivo. Se il presente capitolo ha delineato il "problema" (*l'overload* informativo e le inefficienze del modello tradizionale), il Capitolo 2 esaminerà la "soluzione" tecnologica, analizzando la transizione dai sistemi *rule-based* ai modelli predittivi basati sull'intelligenza artificiale e il suo impatto sull'organizzazione del lavoro del delegato SOS.

Capitolo 2 - L'evoluzione dei modelli di controllo: dall'approccio *Rule-Based* ai sistemi predittivi

Il primo capitolo si è chiuso con una diagnosi: il sistema delle segnalazioni di operazioni sospette funziona, ma mostra segni di affanno. L'eccesso di segnalazioni difensive e l'elevatissima quota di falsi positivi sono sintomi di un modello che fatica a reggere il passo con la realtà che dovrebbe presidiare. Le nuove Istruzioni UIF del dicembre 2025, esaminate nell'ultimo paragrafo, hanno offerto una prima risposta sul piano normativo, aprendo con cautela all'impiego dell'intelligenza artificiale e ribadendo al tempo stesso la centralità del controllo umano. È da qui che prende le mosse il presente capitolo, che costituisce il nucleo dell'intero lavoro.

L'obiettivo è analizzare la transizione in corso nei sistemi di monitoraggio delle transazioni: il passaggio, ancora tutt'altro che concluso, da modelli fondati su regole predefinite a sistemi capaci di apprendere dai dati. Non si tratta di una semplice sostituzione tecnologica, ma di un cambiamento nel modo stesso di concepire l'individuazione dell'anomalia. Dove il modello tradizionale applica regole fisse, il modello predittivo cerca schemi; dove il primo ragiona per soglie, il secondo ragiona per probabilità. Comprendere questa differenza e, soprattutto, le sue implicazioni operative e giuridiche, è essenziale per valutare con equilibrio tanto le ambizioni quanto i rischi della nuova tecnologia.

L'analisi procederà per gradi. Il primo paragrafo ricostruisce i limiti strutturali dei sistemi *rule-based*, così da chiarire le ragioni concrete che spingono il settore verso il cambiamento. Il secondo entra nel merito delle tecniche di intelligenza artificiale e machine learning applicate al *transaction monitoring*, illustrandone il funzionamento e le potenzialità, con particolare attenzione alla capacità di ridurre i falsi positivi e di riconoscere gli schemi di riciclaggio più complessi. Il terzo affronta la dimensione forse più delicata, quella organizzativa: come cambia, in concreto, il lavoro di chi opera nella funzione antiriciclaggio quando l'automazione si fa più sofisticata, e come si ridefinisce il rapporto tra la macchina che elabora e l'uomo che decide. Infine, il quarto introdurrà il cosiddetto problema della "black box".

2.1 Limiti degli schemi rigidi dei sistemi tradizionali

Per comprendere perché il settore antiriciclaggio stia investendo con crescente convinzione nell'intelligenza artificiale, occorre partire da una constatazione: i sistemi di monitoraggio tradizionali, per quanto raffinati nel corso degli anni, scontano alcuni limiti che appaiono ormai strutturali. Non sono malfunzionamenti occasionali, bensì di caratteristiche intrinseche al modo stesso in cui questi sistemi sono concepiti. Averli chiari è il presupposto per valutare, senza facili entusiasmi, cosa la nuova tecnologia possa effettivamente offrire.

Come si è visto nel primo capitolo, procedure come GIANOS operano secondo una logica deterministica, che la dottrina tecnica riassume nella formula "se… allora": il sistema genera un alert al ricorrere di una certa condizione come il superamento di una soglia, una determinata tipologia di operazione oppure una controparte localizzata in un Paese a rischio. Questa impostazione ha il pregio della trasparenza, perché ogni segnalazione è riconducibile a una regola nota e verificabile, ma proprio in tale rigidità risiede la sua principale debolezza[60].

Il primo limite è la staticità. Le regole vengono definite a priori da chi progetta il sistema e restano identiche finché qualcuno non interviene a modificarle manualmente. Il sistema non impara: ripete le stesse valutazioni a prescindere dagli esiti che quelle valutazioni hanno prodotto nel tempo. Se una certa regola genera in modo sistematico allarmi che gli analisti archiviano come infondati, il software continuerà imperterrito a generarli, perché non dispone di alcun meccanismo per apprendere dai propri errori. L'aggiornamento dipende interamente dall'intervento umano, con tutti i ritardi e i costi che questo comporta.

Da qui discende un secondo problema, forse il più insidioso: la prevedibilità del sistema si presta a essere sfruttata da chi intende aggirarlo. Un riciclatore che conosce o intuisce le soglie oltre le quali scatta un controllo, può semplicemente restare al di sotto di esse. È il caso classico dello *smurfing*: il frazionamento di un importo rilevante in molte operazioni di piccola entità, ciascuna delle quali, presa singolarmente, non supera la soglia di attenzione. Paradossalmente, più le regole sono rigide e note, più diventano facili da eludere per chi agisce in malafede: la logica statica finisce per proteggere proprio i comportamenti più sofisticati, mentre intercetta soprattutto quelli casuali[61].

Il terzo limite riguarda la difficoltà di cogliere le relazioni complesse. Un sistema a regole ragiona, per sua natura, sulla singola operazione o sul singolo rapporto: valuta se quella transazione, di per sé, presenta caratteristiche anomale. Fatica invece a ricostruire disegni che si sviluppano attraverso una pluralità di soggetti e di conti collegati, dove nessuna operazione, isolatamente considerata, appare sospetta, ma è l'architettura complessiva a rivelare l'intento illecito. Gli schemi di riciclaggio più evoluti (come le reti di prestanome o le frodi carosello) vivono esattamente in questo spazio, sfuggendo a controlli che non riescono a intercettare l'intera rete di transazioni nel suo complesso[62].

---

[60] Sui limiti dei sistemi di monitoraggio basati su regole, fra cui l'elevato tasso di falsi positivi e la difficoltà di intercettare tecniche di riciclaggio in continua evoluzione, cfr. OZTAS B. – CETINKAYA D. – ADEDOYIN F. – BUDKA M. – AKSU G. – DOGAN H., «Transaction Monitoring in Anti-Money Laundering: A Qualitative Analysis and Points of View from Industry», in *Future Generation Computer Systems*, vol. 159, 2024.

[61] Sul fenomeno dello *smurfing* come tecnica di elusione delle soglie e sui limiti dei controlli a soglia fissa, v. Provvedimento UIF del 12 maggio 2023 recante gli indicatori di anomalia, disponibile su: uif.bancaditalia.it.

[62] Le frodi carosello si basano sull'interposizione di società "cartiere" che acquistano beni da fornitori comunitari in regime di non imponibilità IVA e li rivendono sul mercato nazionale applicando l'imposta, che tuttavia non viene mai versata all'Erario; i relativi proventi vengono poi dispersi attraverso una catena di società-filtro, spesso con destinazione estera. La rilevazione di tali condotte risulta particolarmente ardua per i sistemi tradizionali, poiché ciascuna transazione, isolatamente considerata, appare lecita: è solo la ricostruzione dell'intera rete di soggetti e di flussi a rivelarne la natura fraudolenta. Sul punto v. UIF, *Schemi*

A questi tre limiti se ne aggiunge uno di natura economica, che nel primo capitolo si è già incontrato affrontando il tema dei falsi positivi. La rigidità delle regole produce un volume enorme di allarmi privi di reale fondamento: operatività perfettamente lecite che, per una coincidenza statistica, attivano un criterio di anomalia. Le conseguenze sono duplici. Sul piano dei costi, la funzione antiriciclaggio è costretta a impiegare gran parte delle proprie risorse per smaltire allarmi che si riveleranno infondati, sottraendo tempo e attenzione ai casi che meriterebbero un'analisi approfondita. Sul piano dell'efficacia, l'eccesso di rumore aumenta il rischio che una segnalazione realmente significativa si perda nella massa, secondo un meccanismo che si può e come il cosiddetto "affaticamento da allarme": quando gli avvisi sono troppi e quasi sempre infondati, l'attenzione dell'operatore inevitabilmente si abbassa.

Nessuno di questi limiti, è bene precisarlo, toglie valore al lavoro svolto in questi decenni dai sistemi *rule-based*, che rappresentano l'ossatura del monitoraggio antiriciclaggio. Il punto non è che tali sistemi siano inutili, ma che il contesto in cui operano è cambiato: i volumi di transazioni sono cresciuti a dismisura, le tecniche di riciclaggio si sono fatte più sofisticate e la pressione regolatoria e reputazionale sugli intermediari si è intensificata. In questo scenario, un modello che non apprende, che è facilmente prevedibile e che genera spesso allarmi infondati mostra tutti i segni di uno strumento non conforme all'attualità. È da questa consapevolezza che nasce l'interesse per approcci capaci di superarne i vincoli, di cui ci si occuperà nel prossimo paragrafo.

2.2 Le nuove frontiere del *Transaction Monitoring*: l'uso dell'Intelligenza Artificiale per l'identificazione di anomalie complesse

Quando si parla di "intelligenza artificiale" applicata al monitoraggio delle transazioni, si tende a immaginare un'entità unica capace di intercettare i flussi di denaro "sporco". La realtà è meno suggestiva ma più interessante: l'intelligenza artificiale non è una tecnologia sola, bensì un insieme eterogeneo di tecniche e modelli, ciascuno con logiche di funzionamento, punti di forza e debolezze differenti. Comprendere queste differenze è una condizione necessaria per valutare, con cognizione di causa, cosa questi strumenti possano davvero offrire alla funzione antiriciclaggio e cosa invece resti fuori dalla loro portata.

In termini generali, ciò che accomuna gli approcci di cui si dirà è il ricorso al machine learning, l'apprendimento automatico: una branca dell'intelligenza artificiale in cui il sistema non riceve istruzioni esplicite su cosa cercare, ma ricava autonomamente dai dati i criteri per distinguere ciò che è normale da ciò che è anomalo. È un rovesciamento di prospettiva: non è più il programmatore a stabilire a priori la soglia di sospetto, ma è il modello a costruirsela, imparando dagli esempi che gli vengono forniti. Da questa capacità di apprendere discendono tanto le promesse quanto i rischi della nuova

rappresentativi di comportamenti anomali – Operatività connessa con illeciti fiscali, 10 novembre 2020, disponibile su: uif.bancaditalia.it.

tecnologia, che si cercherà di illustrare senza indulgere né all'entusiasmo acritico né al pregiudizio.

2.2.1 Apprendimento supervisionato e non supervisionato

La prima distinzione da fare è quella tra due modi radicalmente diversi in cui un sistema può imparare dai dati: l'apprendimento supervisionato e l'apprendimento non supervisionato. Non si tratta di una sottigliezza teorica, perché i due approcci rispondono a due esigenze distinte del monitoraggio antiriciclaggio e, come si vedrà, danno il meglio di sé se combinati.

Nell'apprendimento supervisionato (*supervised learning*) il modello viene addestrato su una base di dati storici in cui ogni caso è già stato "etichettato", ossia classificato in partenza come lecito o sospetto. In concreto, si fornisce all'algoritmo un archivio di operazioni passate accompagnate dall'esito che hanno avuto, nello specifico: se queste hanno dato luogo a una segnalazione fondata oppure se si sono rivelate normali. Il sistema analizza questi esempi e, per così dire, ne estrae la "morfologia" del sospetto: individua quali combinazioni di caratteristiche (importi, frequenze, tipologie di controparti, sequenze temporali) ricorrono con maggiore probabilità nei casi realmente illeciti. Una volta addestrato, il modello è in grado di applicare quanto appreso a operazioni nuove, assegnando a ciascuna una probabilità di essere sospetta[63]. Il vantaggio è evidente: il sistema impara dall'esperienza concreta dell'intermediario e in particolare dagli esiti delle segnalazioni passate. Un aspetto che, come si è visto, le nuove Istruzioni UIF valorizzano attraverso il flusso di ritorno sugli esiti delle SOS, che alimenta proprio la capacità di apprendimento del sistema[64].

Questo approccio ha però un limite intrinseco, che ne definisce anche il confine di utilità: il modello può riconoscere soltanto ciò che somiglia a quanto ha già visto. Se una tecnica di riciclaggio è nuova, mai osservata prima e quindi assente dai dati di addestramento, il sistema supervisionato rischia di non coglierla, perché non dispone di esempi da cui averla imparata. È, quindi, molto efficace nel ritrovare schemi noti, ma strutturalmente cieco di fronte all'ignoto.

È qui che entra in gioco l'apprendimento non supervisionato (*unsupervised learning*). In questo caso il modello lavora su dati privi di etichetta: non gli viene detto in anticipo quali operazioni siano lecite e quali sospette, e il suo compito non è classificare, bensì scoprire strutture nascoste nei dati. Il sistema costruisce autonomamente una mappa della "normalità", raggruppando i comportamenti simili in insiemi omogenei (i cosiddetti *cluster*). Esso poi segnala ciò che si discosta in modo significativo da tali insiemi, cioè le anomalie. Non cerca un profilo predefinito di sospetto: cerca semplicemente ciò che è fuori posto, ciò che non assomiglia a nulla di consueto. Questa caratteristica lo rende

[63] Sull'apprendimento supervisionato applicato all'individuazione del riciclaggio e sul confronto tra i principali algoritmi di classificazione, cfr. NGUYEN V.D. – NGUYEN K.H., «Enhancing Money Laundering Detection: A Comparative Study of Machine Learning Techniques and Sampling Methods», in *Computational Science and Its Applications – ICCSA 2025*, Lecture Notes in Computer Science, vol. 15650, Springer, Cham, 2025.

[64] Il riferimento è alla Sezione VIII della Parte Prima delle nuove Istruzioni UIF del 18 dicembre 2025 sul flusso di ritorno relativo agli esiti delle segnalazioni.

prezioso proprio dove il supervisionato fallisce, ossia nell'individuazione di schemi inediti, di comportamenti che nessuno aveva ancora catalogato come illeciti ma che, per la loro stessa stranezza, meritano un approfondimento.

Anche l'approccio non supervisionato ha i suoi costi. Segnalando tutto ciò che è "diverso", tende a produrre un numero elevato di falsi allarmi, perché non tutto ciò che è insolito è anche illecito: un'operazione può essere anomala per mille ragioni perfettamente legittime. La capacità di cogliere l'inatteso è dunque un'arma a doppio taglio, poiché senza il filtro dell'esperienza pregressa fatica a distinguere l'anomalia sospetta da quella innocua.

Proprio perché i due approcci hanno pregi e limiti speculari, nella pratica del monitoraggio antiriciclaggio essi non si escludono, ma si integrano. Un'architettura ben concepita impiega tipicamente il non supervisionato per esplorare i dati e far emergere comportamenti anomali che nessuna regola avrebbe intercettato, e il supervisionato per assegnare a tali comportamenti una priorità di rischio, distinguendo quelli che somigliano a casi già rivelatisi illeciti da quelli probabilmente innocui. Ne risulta un sistema capace, almeno in teoria, di coniugare la scoperta dell'ignoto con il riconoscimento del noto. È su questa combinazione che si innestano i modelli concreti di cui si dirà nel prossimo sotto-paragrafo.

2.2.2 I modelli principali: alberi decisionali, foreste casuali, *gradient boosting* e reti neurali

Ora che è stata chiarita la distinzione di fondo tra apprendimento supervisionato e non supervisionato, si può entrare nel merito dei modelli concretamente impiegati nel monitoraggio delle transazioni. È la parte più tecnica di questa trattazione, e lo è per una ragione precisa: capire, almeno nelle linee essenziali, come questi strumenti arrivino a formulare un giudizio di rischio è indispensabile per discutere in modo consapevole della loro efficacia e dei loro limiti. Si procederà dal modello più semplice e intuitivo verso quelli via via più sofisticati, con l'obiettivo di renderne comprensibile la logica di funzionamento.

Un dato preliminare aiuta a orientarsi. I dati su cui lavora una funzione antiriciclaggio sono in larghissima parte dati strutturati e tabellari: righe che rappresentano operazioni e colonne che ne descrivono le caratteristiche: importo, data, valuta, tipologia, profilo del cliente, conto di provenienza e di destinazione. Su questo genere di dati, per ragioni che si chiariranno, i modelli più efficaci non sono le reti neurali (di cui si approfondirà in seguito), bensì i modelli basati su alberi decisionali. Vale dunque la pena partire proprio da questi.

L'albero decisionale (*decision tree*) è la base di questa famiglia di modelli, ed è anche il più facile da comprendere, perché il suo modo di ragionare assomiglia a quello di un analista che procede per domande successive. L'albero scompone il problema in una sequenza di test binari, ciascuno dei quali divide i dati in due rami: l'importo supera una certa cifra? Sì o no. La controparte è situata in un Paese ad alto rischio? Sì o no. Il conto è stato aperto da meno di sei mesi? E così via, di biforcazione in biforcazione, fino ad arrivare a un punto terminale, la cosiddetta foglia, in cui il modello emette il proprio

verdetto: operazione probabilmente lecita, oppure probabilmente sospetta. La scelta di quali domande porre, e in quale ordine, non è arbitraria: l'algoritmo seleziona a ogni passo il test che separa meglio i casi leciti da quelli illeciti, cercando di rendere ciascun ramo il più omogeneo possibile[65]. Il pregio di questo modello è la trasparenza. Il percorso che porta a una decisione è interamente leggibile e si può ripercorrere all'indietro, spiegando in parole semplici perché una certa operazione sia finita nella foglia "sospetta". È una qualità che, come si vedrà, ha un peso non trascurabile quando si tratta di motivare una segnalazione.

L'albero singolo, però, ha un difetto serio: tende a costruire regole troppo aderenti ai dati su cui è stato addestrato, memorizzandone anche le peculiarità casuali. Il risultato è un modello che funziona benissimo sui casi già visti, ma meno su quelli nuovi, formando un fenomeno noto come *overfitting* (sovradattamento). Per superare questo limite si è pensato di non affidarsi a un albero solo, ma a molti alberi insieme. È l'idea alla base della foresta casuale (*random forest*): anziché costruire un unico albero, se ne costruiscono centinaia, ciascuno addestrato su un sottoinsieme diverso dei dati e su un diverso gruppo di caratteristiche, estratti in modo casuale. Ogni albero esprime il proprio voto sulla singola operazione, e la decisione finale è quella della maggioranza. Il ragionamento è lo stesso che sta dietro alla saggezza di una giuria numerosa e variegata: il singolo giurato può sbagliare, ma l'errore di ciascuno tende a compensare quello degli altri, e il verdetto collettivo risulta più affidabile del giudizio di uno solo[66]. La foresta casuale eredita così buona parte della robustezza che manca all'albero singolo, al prezzo però di una minore leggibilità: seguire il ragionamento di trecento alberi che votano non è la stessa cosa che ripercorrere il percorso di uno singolo.

Un'evoluzione ulteriore, e oggi tra le più diffuse nella pratica, è il *gradient boosting*. Anche qui si combinano molti alberi, ma con una logica diversa e più raffinata. Invece di costruirli tutti in parallelo e farli votare, li si costruisce uno dopo l'altro, in sequenza, chiedendo a ciascun nuovo albero di concentrarsi sugli errori commessi da quelli che lo hanno preceduto. Ogni albero, quindi, non riparte da zero, ma cerca di correggere ciò che l'insieme degli alberi precedenti aveva sbagliato, affinando progressivamente la previsione complessiva. È un procedimento che ricorda il modo in cui si perfeziona una stima per approssimazioni successive, riducendo a ogni passaggio lo scarto rispetto al risultato corretto. L'implementazione più nota di questa tecnica, denominata *XGBoost*, si è imposta come uno degli strumenti più performanti sui dati tabellari e viene infatti frequentemente indicata, negli studi di settore, come il modello che offre i risultati migliori nell'individuazione delle operazioni a rischio di riciclaggio[67]. Questa accuratezza

[65] Per un confronto fra i principali algoritmi di classificazione nell'individuazione del riciclaggio, fra cui i modelli basati su alberi decisionali Random Forest e XGBoost, cfr. AJAGBE S.A. – MAJOLA S. – MUDALI P., «Comparative analysis of machine learning algorithms for money laundering detection», in *Discover Artificial Intelligence*, vol. 5, 2025.

[66] Sull'impiego delle *random forest* nell'individuazione del riciclaggio, cfr. NGUYEN V.D. – NGUYEN K.H., «Enhancing Money Laundering Detection», cit., dove il modello ottiene i risultati migliori fra i sei algoritmi confrontati.

[67] Sulle prestazioni dell'algoritmo *XGBoost*, frequentemente indicato come il più performante sui dati tabellari nell'ambito dell'*anti-money laundering*, cfr. AJAGBE S.A. e altri, «Comparative analysis of machine learning algorithms for money laundering detection», cit., ove il modello riporta i valori più elevati di accuratezza e precisione nel confronto con gli altri algoritmi esaminati.

superiore ha però un rovescio: il modello diventa più difficile da interpretare e, se non calibrato con attenzione, può a sua volta sovradattarsi ai dati.

L'ultima famiglia di modelli è quella delle reti neurali artificiali (*artificial neural networks*). Ispirate, almeno nella metafora, al funzionamento dei neuroni biologici. Queste reti sono composte da unità di calcolo elementari, i cosiddetti neuroni, organizzate in strati successivi. Ogni neurone riceve una serie di segnali in ingresso, li pondera attribuendo a ciascuno un peso, li combina e trasmette un risultato allo strato successivo, dopo averlo filtrato attraverso una funzione che introduce un elemento di non linearità. Semplificando, l'informazione attraversa la rete strato dopo strato, trasformandosi progressivamente. L'apprendimento consiste nel calibrare progressivamente gli innumerevoli pesi che regolano queste connessioni: la rete confronta la propria previsione con l'esito reale, misura lo scarto e lo propaga all'indietro, aggiustando ciascun peso nella direzione che riduce l'errore complessivo. Il processo iterativo si ripete su milioni di esempi finché la previsione non diventa sufficientemente accurata[68]. Quando gli strati sono numerosi si parla di apprendimento profondo (*deep learning*), e la rete acquisisce la capacità di riconoscere autonomamente configurazioni molto complesse, che nessuno ha dovuto descrivere in anticipo. Nel monitoraggio antiriciclaggio, architetture di questo tipo si prestano in particolare ad analizzare le sequenze temporali di operazioni — cogliendo l'evoluzione del comportamento di un conto nel tempo — oppure, in versione non supervisionata, a costruire un modello della normalità e a segnalare ciò che vi si discosta.

Proprio la potenza delle reti neurali, tuttavia, ne costituisce il limite più delicato. A differenza dell'albero decisionale, il cui ragionamento è leggibile passo per passo, una rete neurale profonda distribuisce la propria "conoscenza" su migliaia o milioni di pesi numerici privi di un significato immediatamente interpretabile. Il modello fornisce un risultato: l'operazione è sospetta con una data probabilità, ma senza una spiegazione comprensibile del perché sia giunto a quella conclusione. È il fenomeno che va sotto il nome di "scatola nera" (*black box*): un sistema che decide senza rendere conto del proprio ragionamento in termini accessibili a un essere umano[69]. Su dati strutturati come quelli bancari, peraltro, il vantaggio prestazionale delle reti neurali rispetto ai più trasparenti modelli ad alberi non è sempre così marcato da giustificarne l'opacità, ed è anche per questo che, nella pratica corrente della funzione antiriciclaggio, gli approcci basati su alberi conservano un ruolo di primo piano.

Da questa rassegna emerge un filo da tenere ben presente. Man mano che si sale in sofisticazione (dall'albero singolo alla foresta, dal *gradient boosting* alle reti profonde) cresce la capacità del modello di cogliere schemi complessi, ma diminuisce nella stessa misura la sua trasparenza. La scelta del modello, dunque, non è una questione meramente tecnica rimessa agli informatici: è una scelta che ha ricadute direttamente sulla possibilità di motivare una segnalazione, di controllarla e, in ultima analisi, di attribuirne la

[68] Sulle reti neurali e sulle architetture di apprendimento profondo applicate all'individuazione del riciclaggio, cfr. KUTE D.V. – PRADHAN B. – SHUKLA N. – ALAMRI A., «Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering – A Critical Review», in *IEEE Access*, vol. 9, 2021.

[69] Sul problema della *black box* e sull'esigenza di tecniche di intelligenza artificiale spiegabile (*explainable AI*), cfr. KUTE D.V. e altri, «Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering – A Critical Review», cit.

responsabilità. Un sistema più accurato ma inintelligibile può rivelarsi, paradossalmente, meno utile di uno un po' meno preciso ma spiegabile, perché la funzione antiriciclaggio non deve soltanto individuare l'operazione sospetta, ma anche saper dire, davanti alla UIF e in un'eventuale sede giudiziaria, per quali ragioni la ritiene tale.

2.2.3 L'analisi delle reti: dalla singola operazione alla struttura relazionale

Se c'è un ambito in cui l'intelligenza artificiale mostra un vantaggio netto rispetto ai sistemi tradizionali, è quello dell'analisi delle relazioni tra soggetti. Si è visto nel primo paragrafo di questo capitolo come i sistemi a regole ragionino sulla singola operazione: valutano se quella transazione, presa in sé, superi una soglia o corrisponda a un criterio di anomalia. Molti degli schemi di riciclaggio più insidiosi, però, non sono in una singola operazione anomala, ma nel disegno complessivo che lega tra loro decine o centinaia di operazioni e di soggetti, ciascuno dei quali, isolatamente considerato, appare perfettamente in regola.

La tecnica che affronta il problema da questa angolatura è l'analisi delle reti (*network analysis* o *graph analysis*). L'idea di fondo è semplice: invece di rappresentare i dati come una tabella di operazioni indipendenti, li si rappresenta come un grafo, ossia una struttura fatta di nodi e di collegamenti. I nodi sono i soggetti (i conti, le persone, le imprese) e i collegamenti sono le transazioni che li mettono in relazione. Vista in questo modo, l'operatività di una banca non è più un elenco di righe, ma una mappa di connessioni in cui ogni flusso di denaro traccia un filo conduttore tra due punti. E su una mappa siffatta diventano visibili configurazioni che in una tabella resterebbero invisibili: un soggetto che riceve denaro da moltissime fonti e lo rigira tutto verso un unico destinatario, un gruppo di conti che si scambiano fondi in circolo, una catena di passaggi che allontana progressivamente il denaro dalla sua origine[70].

Si riprendono gli schemi di comportamento anomalo elaborati dalla UIF perché è proprio su di essi che il vantaggio dell'analisi delle reti si apprezza in concreto. Si pensi allo schema dei "muli di denaro": soggetti che ricevono sui propri conti somme di provenienza illecita e le ritrasferiscono rapidamente verso altri conti, trattenendo una piccola percentuale. Guardando il singolo "mulo", l'operatività può non destare particolare allarme; ma osservando il grafo, emerge la struttura tipica a ventaglio (molti conti che convergono verso pochi punti di raccolta) che rivela la rete nel suo insieme. Lo stesso vale per le frodi carosello, in cui una catena di società cartiere si scambia fatture e pagamenti per simulare operazioni commerciali inesistenti: nessuna transazione, da sola, è illecita in modo evidente, ma la ricostruzione della catena di cessioni ne svela la natura fraudolenta. In entrambi i casi, ciò che tradisce il disegno criminale non è il singolo movimento, bensì la topologia della rete, la forma che i flussi assumono quando li si guarda tutti insieme[71].

---

[70] Sull'analisi delle reti (*network analysis*) applicata all'individuazione del riciclaggio e sulla rappresentazione dei flussi finanziari come grafi di nodi e collegamenti, cfr. WEBER M. – DOMENICONI G. – CHEN J. – WEIDELE D.K.I. – BELLEI C. – ROBINSON T. – LEISERSON C.E., «Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics», 2019.

[71] Sugli schemi delle frodi IVA intracomunitarie e del *money muling*, cfr. UIF, *Schemi rappresentativi di comportamenti anomali – Operatività connessa con illeciti fiscali*, 10 novembre 2020, e UIF, *Operatività*

Sul piano tecnico, l'evoluzione più avanzata di questo approccio è rappresentata dalle reti neurali su grafo (*graph neural networks*, in sigla GNN), che applicano la logica dell'apprendimento profondo alle strutture a grafo. Senza addentrarsi in tecnicismi eccessivi, il tratto distintivo di questi modelli è che ogni nodo non viene valutato in isolamento, ma alla luce delle caratteristiche dei nodi a cui è collegato, secondo un meccanismo in cui l'informazione si propaga lungo i collegamenti della rete: ciascun conto, per così dire, viene "descritto" anche attraverso la compagnia che frequenta. È un modo di ragionare che ricalca l'intuizione investigativa secondo cui si conosce un soggetto anche dalle persone con cui si accompagna, e che consente di individuare i nodi sospetti non solo per ciò che fanno, ma per la posizione che occupano nella rete[72].

Le sperimentazioni condotte in questo campo hanno prodotto risultati che meritano di essere letti con attenzione, perché il modo in cui vengono comunemente riferiti si presta a fraintendimenti. Gran parte della ricerca si è concentrata sul dataset Elliptic, un'ampia raccolta pubblica di transazioni in Bitcoin etichettate come lecite o illecite, divenuta il principale banco di prova del settore: lo studio che per primo vi ha applicato in modo sistematico le reti neurali su grafo ha aperto un filone poi ampiamente sviluppato[73].

Un risultato di quello studio va segnalato subito, perché smentisce un'aspettativa diffusa. Nel confronto fra i diversi modelli le reti neurali su grafo non hanno prevalso: a ottenere le prestazioni migliori è stato il *random forest*, cioè uno dei modelli ad alberi descritti in precedenza, che ha superato tanto la regressione logistica quanto le architetture a grafo[74]. Il contributo specifico di queste ultime si è manifestato altrove, nella capacità di sfruttare l'informazione relazionale e di mantenere prestazioni più stabili di fronte al mutare delle condizioni nel tempo. La lezione, per il discorso che si sta conducendo, è che la sofisticazione del modello non si traduce automaticamente in una maggiore capacità di individuazione, e che i modelli più leggibili restano competitivi perfino in un ambito, come quello relazionale, nel quale ci si aspetterebbe il contrario.

Due ultime considerazioni. La prima è che i risultati richiamati provengono in larga misura da ambienti sperimentali e da un dataset specifico, quello delle criptovalute, dove la tracciabilità pubblica della blockchain offre una mole di dati relazionali particolarmente favorevole a questo tipo di modelli: la loro trasposizione all'operatività bancaria tradizionale, con dati più eterogenei ed esigenze di integrazione con i sistemi esistenti, non è immediata né scontata. La seconda, più rilevante ai fini di questa tesi, riguarda la trasparenza. Individuare una rete sospetta è soltanto metà del compito: l'altra metà consiste nel saper spiegare perché quella rete è sospetta, in termini comprensibili a un analista, a un'autorità di vigilanza e, se necessario, a un giudice.

---

connessa con truffe, frodi agevolate dalla tecnologia, money muling e altri reati informatici, Comunicazione dell'8 giugno 2026, entrambi disponibili su: uif.bancaditalia.it.

[72] Sul funzionamento delle reti neurali su grafo (*graph neural networks*) e sul meccanismo di propagazione dell'informazione tra nodi collegati, cfr. WEBER M. e altri, «Anti-Money Laundering in Bitcoin».

[73] Il riferimento è allo studio di WEBER M. e altri, cit., condotto sul dataset Elliptic, ampia raccolta pubblica di transazioni in Bitcoin etichettate come lecite o illecite, divenuta il principale banco di prova per la ricerca sull'applicazione dei modelli a grafo all'*anti-money laundering*.

[74] WEBER M. e altri, «Anti-Money Laundering in Bitcoin», cit., ove nel confronto sperimentale il modello *Random Forest* consegue i punteggi F1 più elevati, superando sia la regressione logistica sia le architetture *Graph Convolutional Network*.

2.2.4 I risultati concreti: benefici, diffusione effettiva e limiti

Adesso bisogna valutare se tutto questo, nella pratica, possa funzionare. La risposta che si ricava dagli studi e dall'esperienza sul campo è sì, ma con riserva, e a due condizioni: che ci si metta d'accordo su cosa voglia dire "funzionare" e che i numeri in circolazione si leggano con una certa diffidenza. Ai modelli di apprendimento automatico si attribuiscono di solito tre vantaggi, ed è su questi che conviene ragionare: meno falsi positivi, analisi più rapida, capacità di adattarsi nel tempo.

Il primo e più decantato vantaggio riguarda la riduzione dei falsi positivi, ossia proprio quella zavorra che, come si è visto nel primo capitolo, affligge i sistemi tradizionali con tassi percentuali notevoli. Su questo terreno le aspettative sono elevate. Gli operatori del settore riferiscono, in relazione a implementazioni presso singoli intermediari, riduzioni dei falsi positivi dell'ordine del cinquanta-sessanta per cento rispetto ai preesistenti sistemi *rule-based*, con punte anche superiori in alcuni casi[75]. Questi dati vanno però maneggiati con cautela, per due ragioni. La prima è che provengono in larga parte dai fornitori delle stesse soluzioni tecnologiche, e risentono quindi di un evidente interesse promozionale; la seconda è che si riferiscono a contesti operativi specifici, difficilmente comparabili tra loro e non sempre trasparenti nella metodologia di misurazione. Ciò non toglie che la direzione sia chiara e ampiamente confermata: la capacità dei modelli di apprendere dai dati storici consente di riconoscere come normali quelle operatività che, pur superando le soglie rigide, presentano caratteristiche ricorrenti di legittimità per quello specifico cliente. È il meccanismo della segmentazione comportamentale, di cui si dirà tra breve, a rendere possibile questo affinamento.

Il beneficio, va sottolineato, non è meramente quantitativo. Ridurre il numero di falsi allarmi significa liberare risorse analitiche e restituire tempo qualificato all'esame dei casi realmente rischiosi. Il valore dell'intelligenza artificiale, in questa prospettiva, non sta tanto nel sostituire l'analista quanto nel sottrarlo a un lavoro ripetitivo e a basso valore aggiunto, restituendogli la possibilità di concentrarsi sull'attività investigativa vera e propria.

Il secondo ordine di benefici attiene alla tempestività. I sistemi tradizionali operano tipicamente in modalità differita, elaborando i dati a intervalli (tipicamente notturni o comunque periodici) e generando gli alert a posteriori, quando l'operazione è già stata eseguita. I modelli di apprendimento automatico, per contro, si prestano a un'analisi in tempo reale o quasi, in grado di valutare l'operazione nel momento stesso in cui viene disposta. Questa capacità apre delle possibilità operative significative, prima fra tutte quella di intercettare un'operazione sospetta prima della sua esecuzione, dischiudendo margini di intervento che con l'analisi differita sarebbero preclusi. La rilevazione in tempo reale è tanto più considerevole nel contesto dei pagamenti istantanei, ormai diffusi, dove la finestra temporale per un eventuale blocco si misura in secondi.

[75] Le stime relative alla riduzione dei falsi positivi nell'ordine del 50-60% sono riferite da operatori e fornitori di soluzioni tecnologiche per l'*anti-money laundering*, sulla base di implementazioni presso singoli intermediari. Trattandosi di dati diffusi dagli stessi fornitori, essi vanno assunti con cautela e a titolo indicativo.

Il terzo beneficio, forse il più caratterizzante, è la capacità di adattamento. Un sistema a regole è per definizione statico: continua ad applicare i medesimi criteri finché un intervento umano non li modifica. Un modello di apprendimento automatico, se opportunamente riaddestrato, può invece aggiornare la propria nozione di "normalità" con l'evolversi dei comportamenti, cogliendo la deriva progressiva delle tecniche di riciclaggio. È qui che si colloca la già menzionata segmentazione comportamentale adattiva: anziché applicare a tutti i clienti le medesime soglie, il sistema costruisce profili di normalità differenziati e li aggiorna nel tempo, valutando ciascuna operazione non rispetto a un parametro assoluto, ma rispetto al comportamento abituale di quello specifico soggetto e del gruppo omogeneo cui appartiene. Un movimento che sarebbe anomalo per un pensionato può essere del tutto fisiologico per un imprenditore, e viceversa: la capacità di tener conto di questa relatività è precisamente ciò che manca ai sistemi a soglia fissa.

I benefici descritti sono reali, ma si accompagnano a costi e a limiti non trascurabili. L'adozione di questi sistemi richiede investimenti ingenti, competenze specialistiche non facilmente reperibili e, soprattutto, grandi quantità di dati storici di qualità su cui addestrare i modelli, una condizione che le realtà di minori dimensioni faticano a soddisfare. Vi è poi la questione, già più volte affiorata, della trasparenza: i modelli più efficaci nel ridurre i falsi positivi sono spesso i meno interpretabili, e questo genera una tensione tra l'efficienza operativa e la capacità di rendere conto delle proprie decisioni. Infine, la capacità di adattamento porta con sé un rischio speculare, quello che il modello apprenda distorsioni presenti nei dati o si adatti in modo improprio, producendo discriminazioni o errori sistematici difficili da individuare proprio a causa dell'opacità del sistema[76].

Questi limiti appena descritti pesano concretamente, e lo conferma un'indagine condotta dall'Unità di Supervisione e Normativa Antiriciclaggio (SNA) della Banca d'Italia su un campione di banche medie, selezionate proprio perché avevano mostrato un'elevata propensione all'innovazione, i cui esiti sono stati pubblicati nel settembre 2025[77]. Se nell'acquisizione della clientela la diffusione degli strumenti innovativi è ormai consolidata, per il monitoraggio transazionale soltanto un quinto delle banche del campione ha adottato tecnologie innovative, e spesso le analisi avvengono ancora tramite processi e riscontri caratterizzati da ambiti di manualità[78]. Nelle tabelle che accompagnano l'indagine, il grado di diffusione delle tecniche di machine learning per migliorare la capacità di individuazione delle anomalie è classificato come basso. Vale la pena sottolineare che si tratta di banche già orientate all'innovazione: il dato, riferito all'intero sistema, sarebbe verosimilmente ancora più contenuto.

---

[76] Sui rischi connessi all'opacità dei sistemi e al loro impiego quando la logica è predeterminata da soggetti esterni all'intermediario, cfr. LEMMA V., «Intelligenza Artificiale e sistemi di controllo: quali prospettive regolamentari?», in *Rivista Trimestrale di Diritto dell'Economia*, Supplemento n. 2 al n. 3/2021.

[77] Banca d'Italia, *Indagine qualitativa sull'adozione di strumenti innovativi per l'adempimento degli obblighi AML/CFT*, 2 settembre 2025, disponibile su: bancaditalia.it. L'indagine è stata condotta dall'Unità di Supervisione e normativa antiriciclaggio (SNA) su un campione di banche medie individuate, sulla base delle indagini Fintech dell'Istituto, tra quelle caratterizzate da un'elevata propensione all'innovazione.

[78] Banca d'Italia, *Indagine qualitativa*, cit., par. 2, lett. a), e Tabella 2, ove il grado di diffusione dell'intelligenza artificiale per il monitoraggio transazionale è classificato come basso.

Il medesimo documento, tuttavia, registra un'aspettativa unanime: tutti gli intervistati convengono che proprio nel monitoraggio transazionale potrebbero registrarsi in prospettiva i maggiori sviluppi, trattandosi di un'attività che richiede controlli su una vasta e crescente mole di dati e informazioni, anche di origine eterogenea[79]. Si delinea dunque uno scarto interessante tra il presente e l'atteso: l'intelligenza artificiale è oggi poco diffusa proprio là dove tutti la ritengono più promettente. È una fotografia che invita a collocare correttamente il discorso di questo capitolo, che riguarda non tanto ciò che le banche italiane fanno oggi, quanto ciò verso cui si stanno muovendo.

L'indagine conferma peraltro, e da fonte istituzionale, i benefici sin qui descritti. Le tecniche di *big data*, *advanced analytics* e *machine learning* consentono di aggiornare il profilo del cliente nel continuo sfruttando un ampio insieme di fonti interne ed esterne; di individuare i soggetti correlati a quelli considerati a rischio tramite relazioni statiche (come cointestazioni e cariche societarie) e dinamiche (come le operazioni di trasferimento di denaro); e di calcolare il livello di rischio dei singoli clienti e del rispettivo *cluster* sulla base di algoritmi probabilistici che tengono conto del loro modello di business[80]. Quanto ai falsi positivi, la valutazione della Banca d'Italia è netta: la maggior precisione delle analisi permette di migliorare l'accuratezza dei risultati e di ridurre i falsi positivi, ottimizzando i controlli di secondo livello. Si ritrovano qui, riformulati nel linguaggio della vigilanza, tanto l'analisi delle reti quanto la segmentazione per *cluster* di cui si è detto.

Vi è però un'ulteriore precisazione, contenuta nella medesima indagine, che merita di essere valorizzata perché avvalora la cautela con cui si sono sin qui trattati i dati sulle prestazioni. Nei processi di monitoraggio transazionale, osserva la Banca d'Italia, le evidenze raccolte presso il campione sono ancora parziali, poiché l'innovazione tecnologica in queste fasi è avvenuta più di recente, e spesso in parallelo e a integrazione delle soluzioni tradizionali; ciò non consente di raccogliere dati sufficienti ad apprezzare in modo statisticamente significativo le maggiori prestazioni rappresentate dalle banche stesse[81]. È una conferma autorevole del fatto che le percentuali che circolano in materia vanno assunte come indicazioni di tendenza e non come misure consolidate.

Tornando ai limiti, la stessa indagine della Banca d'Italia individua con precisione le aree di rischio connesse all'impiego di strumenti innovativi in ambito antiriciclaggio: i rischi informatici, quelli di frode, quelli legati all'esternalizzazione e quelli legati alla mancata comprensione dei meccanismi di funzionamento dei sistemi impiegati. L'autorità di vigilanza, in altre parole, riconosce esplicitamente il problema dell'opacità. Osserva inoltre che il maggior ricorso a fornitori esterni specializzati, se da un lato permette di far leva su competenze specialistiche e sistemi in costante aggiornamento, dall'altro può rendere difficoltosi i controlli su alcune fasi essenziali dei processi, tra cui proprio il monitoraggio dell'operatività. Vi è infine un dato che dovrebbe far riflettere chi confida in un progresso automatico: dall'indagine emerge che alcuni intermediari hanno rinunciato a progetti già avviati, perché le nuove soluzioni, introdotte senza un adeguato

[79] Banca d'Italia, *Indagine qualitativa*, cit., par. 2, lett. a).
[80] Banca d'Italia, *Indagine qualitativa*, cit., par. 2, lett. b).
[81] Banca d'Italia, *Indagine qualitativa*, cit., par. 2, lett. b), in ordine alla parzialità delle evidenze raccolte sul monitoraggio transazionale.

vaglio critico, non hanno dispiegato i benefici promessi rispetto agli applicativi tradizionali[82].

In definitiva, i risultati concreti confermano che l'intelligenza artificiale offre risposte serie alle inefficienze strutturali del modello tradizionale, ma non costituisce la soluzione automatica e priva di controindicazioni. È uno strumento potente che, più che eliminare il problema, lo sposta (dalla quantità degli alert alla qualità e alla governabilità del processo): dove prima l'intermediario doveva gestire l'inefficienza dei falsi positivi, ora deve gestire la complessità, l'opacità e i costi di sistemi molto più sofisticati.

2.2.5 La soglia di segnalazione: ciò che la tecnologia non cambia

I benefici descritti nel sotto-paragrafo precedente riguardano la capacità del sistema di distinguere le operazioni sospette da quelle lecite. Esiste però un secondo elemento che concorre a determinare quante segnalazioni un intermediario produce, e che nessun modello di apprendimento automatico tocca: la soglia oltre la quale l'intermediario decide di segnalare. È un aspetto che si esamina con gli strumenti dell'analisi economica, perché permette di collegare il problema del *defensive reporting*, esaminato nel primo capitolo, alle promesse della tecnologia.

Si consideri un'operazione alla quale il sistema attribuisce una probabilità p di essere illecita. Se l'intermediario la segnala, sostiene un costo c, costituito dal tempo di analisi e dagli oneri di compilazione, sostanzialmente indipendente dall'esito. Se non la segnala e l'operazione si rivela illecita, si espone con probabilità q all'accertamento dell'omissione e, in quel caso, a una sanzione F e a un danno reputazionale R. Un intermediario razionale segnala quando il costo atteso dell'omissione supera quello della segnalazione, ossia quando

$$p \cdot q \cdot (F + R) > c$$

e dunque quando la probabilità attribuita all'operazione supera la soglia

$$t = c / [q \cdot (F + R)]$$

La soglia, in altre parole, non dipende dal modello, ma dal rapporto fra i costi[83]. E poiché il costo di una segnalazione è modesto mentre quello di un'omissione accertata può essere assai elevato, la soglia che ne risulta è bassa: l'intermediario ha convenienza a segnalare anche operazioni cui attribuisce una probabilità di illiceità contenuta. È la formalizzazione, in termini semplificati, dell'asimmetria di incentivi descritta dalla letteratura sul gridare «al lupo»[84].

---

[82] Banca d'Italia, *Indagine qualitativa*, cit., par. 2, lett. b), in ordine ai rischi ICT, di frode e di esternalizzazione, nonché a quelli connessi alla mancata comprensione dei meccanismi di funzionamento dei sistemi impiegati, e alla rinuncia a progetti già avviati da parte di alcuni intermediari.

[83] Il modello costituisce elaborazione propria dell'autore, costruita in forma semplificata a fini illustrativi sulla base dell'impostazione di TAKÁTS E., «A Theory of "Crying Wolf"», cit.

[84] TAKÁTS E., «A Theory of "Crying Wolf"», cit.; per il caso italiano cfr. DALLA PELLEGRINA L. – DI MAIO G. – MASCIANDARO D. – SARACENO M., «Are Bankers "Crying Wolf"?», cit.

Cosa cambia, in questo schema, l'adozione di un modello di apprendimento automatico? Il suo effetto principale riguarda la qualità della stima di *p*. Un modello più accurato separa meglio le due popolazioni, attribuendo probabilità più elevate alle operazioni effettivamente illecite e più basse a quelle lecite: a parità di soglia, un numero minore di operazioni lecite la supera, e i falsi positivi diminuiscono. È il beneficio atteso, ed è reale. La sua entità, tuttavia, dipende dalla posizione della soglia. Quando questa è molto bassa, anche operazioni lecite a cui il modello attribuisce probabilità ridotte, ma non nulle, continuano a superarla, e il guadagno si comprime.

Vi è poi un secondo effetto, meno evidente e di segno opposto. La tecnologia non migliora soltanto la stima del rischio: riduce anche il costo dell'analisi, perché gli alert arrivano all'operatore già ordinati per priorità e corredati di elementi informativi. Una riduzione di *c*, nella formula, abbassa ulteriormente la soglia. Ne deriva un possibile paradosso: l'efficienza guadagnata nella lavorazione degli alert può essere in parte riassorbita da un aumento delle segnalazioni, secondo un meccanismo analogo all'effetto rimbalzo che l'economia dell'energia conosce da tempo, per cui un miglioramento dell'efficienza tende ad accrescere l'uso della risorsa che si intendeva risparmiare.

Lo schema ha evidenti limiti: assume che le probabilità stimate dal modello siano correttamente calibrate, che l'intermediario sia neutrale al rischio e che il costo della segnalazione sia costante. Consente però di fissare un punto che attraversa l'intera trattazione. La tecnologia agisce sulla qualità della stima; gli incentivi agiscono sulla soglia. Se questi ultimi restano asimmetrici, un sistema di monitoraggio più preciso migliora la selezione ma non elimina la propensione a segnalare nel dubbio.

Le leve per intervenire sulla soglia sono dunque diverse da quelle tecnologiche, e appartengono all'architettura regolatoria. La letteratura economica ha suggerito di calibrare le sanzioni o di introdurre un costo per la segnalazione[85]. Il regolatore italiano ha scelto un'altra strada, che agisce direttamente sul comportamento: le Istruzioni della UIF del dicembre 2025, escludendo che il superamento di soglie quantitative, il rischio elevato attribuito al cliente o la ricezione di richieste delle autorità possano di per sé fondare una segnalazione, tentano precisamente di innalzare per via normativa la soglia che gli incentivi tenderebbero ad abbassare[86]. Allo stesso modo, il flusso di ritorno sugli esiti delle segnalazioni, fornendo all'intermediario indicazioni su quali comunicazioni siano risultate utili, può modificarne la percezione del rischio di omissione. Sono interventi che operano sul lato degli incentivi e che, proprio per questo, costituiscono il complemento necessario, e non un'alternativa, all'adozione di strumenti tecnologici più evoluti.

[85] TAKÁTS E., «A Theory of "Crying Wolf"», cit.
[86] UIF – Unità di Informazione Finanziaria per l'Italia, *Istruzioni per la rilevazione e la segnalazione delle operazioni sospette*, Provvedimento del 18 dicembre 2025, Parte Prima, Sezioni I e IV, disponibile su: uif.bancaditalia.it.

2.2.6 Caso studio della UIF: un modello di *machine learning* per l'identificazione di aziende collegate alla criminalità organizzata in Italia

Il discorso condotto sin qui si è mosso, per necessità, su un piano prevalentemente teorico: si sono illustrate tecniche, potenzialità e limiti facendo riferimento alla letteratura e a evidenze di settore. Vale però la pena chiudere il paragrafo con un caso concreto, e per una ragione che va oltre l'esemplificazione. L'esperienza che si intende richiamare non proviene infatti da un intermediario né da un fornitore di soluzioni tecnologiche, ma dall'autorità che di quelle segnalazioni è destinataria: la stessa Unità di Informazione Finanziaria ha sperimentato un modello di apprendimento automatico, dandone conto in uno studio pubblicato nel giugno 2024[87].

L'obiettivo del lavoro è l'identificazione delle imprese potenzialmente collegate alla criminalità organizzata: un tema che si colloca al cuore delle finalità del sistema di prevenzione, giacché l'infiltrazione delle organizzazioni criminali nell'economia legale passa in larga misura attraverso strutture societarie apparentemente regolari. L'impianto è quello dell'apprendimento supervisionato descritto in apertura di questo paragrafo. Il modello è stato costruito su un insieme di dati relativi a imprese italiane, ottenuto integrando informazioni finanziarie provenienti da varie fonti, tra cui principalmente dati di bilancio; per l'addestramento e la verifica è stato utilizzato un campione di oltre ventottomila aziende caratterizzate da un'elevata probabilità di essere collegate alla criminalità organizzata, posto a confronto con sottoinsiemi di aziende presumibilmente sane selezionati in modo casuale[88]. Si tratta, in sostanza, di insegnare all'algoritmo a distinguere due popolazioni sulla base delle tracce che l'infiltrazione criminale lascia nei conti di un'impresa.

I risultati riportati sono di un certo interesse. In fase di test, l'algoritmo identifica con successo circa il settantasei per cento delle aziende collegate alla criminalità organizzata e il settantaquattro per cento di quelle presumibilmente sane[89]. Le due misure, che nel linguaggio tecnico prendono il nome di *recall* e *specificity*, esprimono aspetti diversi e complementari della prestazione: la prima indica quanta parte dei casi effettivamente illeciti il modello riesce a intercettare, la seconda quanta parte dei casi leciti riesce a lasciare correttamente fuori. La loro lettura congiunta è essenziale, perché un modello può facilmente eccellere nell'una a scapito dell'altra: segnalando tutto, si intercetterebbero tutti gli illeciti, al prezzo di travolgere anche gli innocenti. L'equilibrio tra i due valori, poco sopra i tre quarti in entrambi i casi, restituisce l'immagine di uno strumento utile ma tutt'altro che infallibile: circa un'impresa criminale su quattro sfugge al modello, e circa un'impresa sana su quattro viene erroneamente attenzionata.

---

[87] UIF – Unità di Informazione Finanziaria per l'Italia, CARIELLO P. – DE SIMONI M. – IEZZI S., «Un modello di *machine learning* per l'identificazione di aziende collegate alla criminalità organizzata in Italia», in *Quaderni dell'antiriciclaggio – Analisi e studi*, n. 22, 10 giugno 2024, disponibile su: uif.bancaditalia.it. Lo studio è pubblicato per esteso in lingua inglese con il titolo «A machine learning approach for the detection of firms linked to organised crime in Italy», corredato di un estratto non tecnico in italiano.

[88] CARIELLO P. – DE SIMONI M. – IEZZI S., «Un modello di *machine learning*», cit.

[89] CARIELLO P. – DE SIMONI M. – IEZZI S., «Un modello di *machine learning*», cit., ove i risultati sono espressi in termini di *recall* (76%) e *specificity* (74%).

Vale la pena tradurre questi valori in termini operativi, perché il risultato è meno rassicurante di quanto le due percentuali suggeriscano. Si ipotizzi, a fini illustrativi, che le imprese effettivamente collegate alla criminalità organizzata rappresentino il due per cento del totale, e si applichi il modello a centomila imprese. Esso individuerebbe circa millecinquecento delle duemila imprese criminali, ma segnalerebbe al tempo stesso oltre venticinquemila imprese sane: su ventisettemila posizioni attenzionate, quelle effettivamente rilevanti sarebbero meno del sei per cento. Il dato non sminuisce il valore dello strumento, ma ne chiarisce la natura: un sistema del genere non individua le imprese infiltrate, bensì restringe drasticamente il campo entro cui cercarle. Il che conferma, per altra via, la qualificazione che gli stessi autori ne danno.

Ed è proprio qui che il caso diventa istruttivo per il discorso di questa tesi. Il principale prodotto dell'algoritmo, spiega lo studio, è un punteggio di rischio, che potrebbe essere utilizzato a livello operativo per supportare l'azione delle autorità antiriciclaggio e delle forze dell'ordine, ad esempio come strumento di *screening* preliminare[90]. La formula usata è misurata e va letta con attenzione: si parla di supporto e di *screening* preliminare, non di individuazione né tantomeno di accertamento. Il modello non stabilisce che un'impresa sia infiltrata; ordina un insieme troppo vasto per essere esaminato per intero, indicando da dove cominciare. La decisione su cosa quel punteggio significhi resta interamente altrove.

Non meno significativa è la chiusura dello studio, che merita di essere sottolineata perché proviene da chi il modello lo ha costruito: conferme della sua validità operativa dovranno venire da ulteriori applicazioni sul campo[91]. È un'affermazione di prudenza scientifica che conferma quanto osservato poco sopra a proposito della non definitività delle evidenze disponibili. Chi ha sviluppato l'algoritmo si guarda bene dal presentarlo come una soluzione: lo propone come ipotesi da verificare nella pratica.

Il caso, in definitiva, offre una conferma su due piani. Sul piano tecnico, mostra che l'apprendimento automatico funziona davvero su problemi reali dell'antiriciclaggio, e che lo fa con risultati apprezzabili benché lontani dalla perfezione. Sul piano concettuale, mostra che perfino l'autorità che lo ha sviluppato lo colloca dove questa trattazione sostiene debba stare: a monte del giudizio, come strumento che seleziona e ordina il materiale su cui altri (analisti, investigatori, magistrati) eserciteranno la propria valutazione. La macchina restringe il campo, la decisione resta umana. È una simmetria interessante, peraltro, rispetto a quanto avviene presso gli intermediari: tanto chi le segnalazioni le produce quanto chi le riceve si sta attrezzando con strumenti analoghi, e in entrambi i casi il problema che si pone non è se la tecnologia funzioni, ma come governarne il rapporto con chi deve rispondere delle conclusioni.

[90] CARIELLO P. – DE SIMONI M. – IEZZI S., «Un modello di *machine learning*», cit.
[91] CARIELLO P. – DE SIMONI M. – IEZZI S., «Un modello di *machine learning*», cit.

2.3 L'impatto organizzativo: come cambia il lavoro del delegato SOS con il supporto di strumenti tecnologici evoluti (relazione *System-to-Human*)

I due paragrafi precedenti hanno guardato all'intelligenza artificiale dal punto di vista degli strumenti: come funzionano i modelli, cosa sanno fare e quali risultati promettono. Ma un sistema di monitoraggio, per quanto sofisticato, non opera nel vuoto: si inserisce in un'organizzazione fatta di persone, ruoli, responsabilità e procedure. Ed è su questo tessuto organizzativo che l'adozione dell'intelligenza artificiale produce alcuni degli effetti più profondi e anche più sottovalutati. Chiudendo il paragrafo precedente si è osservato come la tecnologia, più che eliminare i problemi, li sposti: dalla gestione della quantità di *alert* alla governabilità di un processo diventato più complesso. È esattamente questo spostamento a ridisegnare il lavoro di chi, in banca, si occupa di antiriciclaggio, e in particolare la posizione del delegato alle segnalazioni di operazioni sospette.

Conviene allora cambiare prospettiva e chiedersi non più cosa faccia la macchina, ma cosa resti (e cosa cambi) per l'uomo che le sta accanto. La questione non è secondaria né meramente descrittiva: come si vedrà, il modo in cui si configura il rapporto tra il sistema automatizzato e il decisore umano ha implicazioni dirette sulla tenuta stessa del sistema di prevenzione, perché è da quel rapporto che dipende chi, in ultima analisi, risponde di una segnalazione inviata o omessa. Il paragrafo affronta il tema in due passaggi: dapprima osservando come mutano concretamente il ruolo dell'analista e del delegato quando il monitoraggio si fa intelligente; poi inquadrando sul piano concettuale e normativo la relazione tra sistema e uomo, quella che la dottrina definisce relazione *system-to-human*.

2.3.1 Dal filtro degli *alert* all'analisi qualificata: la trasformazione del ruolo

Per cogliere la portata del cambiamento, si può partire da come si svolge, oggi, il lavoro di un analista antiriciclaggio in un intermediario che si affida ai sistemi tradizionali. Si è visto nel primo capitolo che questi sistemi generano un volume elevatissimo di *alert*, dei quali la grande maggioranza si rivela infondata. Ne consegue che una porzione rilevante del tempo dell'analista viene assorbita dalla verifica di allarmi che si riveleranno infondati. Escludere un falso positivo richiede competenza e tempo, perché occorre comprendere l'operazione, ricostruirne il contesto e motivare l'archiviazione. È un'attività sproporzionata rispetto al suo effettivo rendimento: a fronte di un impegno considerevole, solo una piccola frazione delle verifiche si conclude con una segnalazione. Sorge quindi il rischio che le poche situazioni realmente rischiose ricevano meno attenzione di quanta ne meriterebbero, disperse in un volume di lavoro ampio e, spesso, sterile.

L'introduzione di sistemi capaci di ridurre in modo significativo i falsi positivi modifica alla radice questo scenario. Se la macchina filtra meglio, l'analista riceve un numero minore di segnalazioni, ma mediamente più pertinenti. Il suo lavoro si sposta così dal versante quantitativo a quello qualitativo: meno tempo speso a scartare allarmi inutili, più tempo dedicato all'approfondimento dei casi che presentano un reale sospetto di rischio. In termini concreti, l'attività muta natura, passando da un compito di smistamento (vagliare, scartare, archiviare) a un compito di investigazione (ricostruire, interpretare,

valutare). È una trasformazione che eleva il profilo professionale richiesto: dove prima bastava applicare con diligenza una procedura di screening, ora serve una capacità di analisi più sofisticata, in grado di leggere schemi complessi e di dialogare criticamente con gli output del sistema.

Sarebbe però un errore leggere questo passaggio come una progressiva marginalizzazione dell'elemento umano, quasi che la macchina, facendo sempre di più, lasci all'uomo sempre di meno. È vero l'opposto. Man mano che il sistema si occupa del lavoro ripetitivo, all'analista resta la parte più delicata e meno automatizzabile: il giudizio. Come si è osservato fin dal primo capitolo, vi sono aspetti qualitativi e relazionali — il contesto di un'operazione, la coerenza di una giustificazione, la conoscenza diretta del cliente — che sfuggono per loro natura all'analisi algoritmica e che solo una valutazione umana può cogliere. L'intelligenza artificiale, per quanto raffinata, produce indizi statistici, non certezze; segnala correlazioni, non ne comprende il significato. Spetta all'analista, e in ultima istanza al delegato, trasformare quell'indizio in una valutazione motivata, calandolo nel contesto specifico e verificandone la tenuta.

Ne consegue che, l'automazione, anziché ridurre l'importanza del fattore umano, ne accresce il peso specifico: sollevato dal lavoro meccanico, l'analista è chiamato a esercitare proprio quelle competenze di giudizio, interpretazione e contestualizzazione che rappresentano il vero valore aggiunto della funzione antiriciclaggio, e che nessun modello, al momento, è in grado di replicare. Il rischio, semmai, è di segno opposto a quello temuto: non che la macchina renda superfluo l'uomo, ma che l'uomo, fidandosi troppo della macchina, abdichi al proprio ruolo di controllo, adagiandosi acriticamente sugli output del sistema. È un rischio — quello del cosiddetto *automation bias*, la tendenza a dare per buono ciò che il sistema suggerisce — che il prossimo sotto-paragrafo affronterà nel quadro più generale della relazione tra sistema e decisore umano.

2.3.2 La relazione *System-to-Human*: automazione, giudizio e responsabilità

Il mutamento del ruolo dell'analista, descritto nel sotto-paragrafo precedente, è la manifestazione concreta di una questione più profonda, che investe il modo stesso in cui l'intelligenza artificiale si inserisce nell'organizzazione dell'intermediario e si rapporta all'intelligenza umana. Per inquadrarla si ricorre a una distinzione che la dottrina ha impiegato con efficacia e che offre una chiave di lettura preziosa: quella tra relazioni *system-to-system* (S2S) e relazioni *system-to-human* (S2H).

Da un lato si collocano gli ecosistemi propriamente artificiali, nei quali l'output di un sistema automatizzato si scambia con un altro sistema, senza intervento umano (S2S); dall'altro i contesti misti, nei quali il risultato prodotto dall'intelligenza artificiale si confronta con l'intelligenza naturale di un essere umano (S2H)[92]. Nel primo caso ciò che si rileva è essenzialmente la rispondenza degli algoritmi a criteri di efficienza; nel secondo si aprono questioni ulteriori, riconducibili alle differenze e alle somiglianze tra il

[92] La distinzione tra relazioni *system-to-system* (S2S) e relazioni *system-to-human* (S2H) è impiegata, con specifico riferimento ai rapporti tra intelligenza artificiale e sistemi di controllo nell'ordinamento bancario e finanziario, da LEMMA V., «Intelligenza Artificiale e sistemi di controllo: quali prospettive regolamentari?», in *Rivista Trimestrale di Diritto dell'Economia*, Supplemento n. 2 al n. 3/2021.

ragionamento della macchina e quello dell'uomo[93]. La funzione antiriciclaggio, per come è stata descritta, è un tipico ambiente *system-to-human*: il sistema di monitoraggio elabora i dati e produce un'evidenza (un *alert*, un punteggio di rischio o una rete di soggetti sospetti), ma quell'evidenza non si traduce automaticamente in una decisione; viene consegnata a un operatore che la valuta, la interpreta e, se del caso, la trasforma in una segnalazione.

Questa collocazione non è un dettaglio tecnico, ma il cuore giuridico della questione. In un rapporto *system-to-human* la macchina non decide: prepara la decisione. Il momento genuinamente decisionale — quello che impegna la responsabilità dell'intermediario e del delegato — resta appannaggio dell'uomo. È una ripartizione che risponde a una precisa logica di sistema: l'ordinamento antiriciclaggio attribuisce l'obbligo di segnalazione a soggetti determinati e ne fa discendere un articolato regime di responsabilità, di natura amministrativa quanto all'omessa segnalazione e potenzialmente penale ove la condotta integri autonome fattispecie di reato. Queste responsabilità presuppongono un centro di imputazione umano, capace di rispondere delle proprie scelte; un algoritmo, allo stato del diritto, non è un soggetto giuridico e non può essere destinatario di un rimprovero. Delegare per intero alla macchina la decisione di segnalare, o di non segnalare, significherebbe creare un vuoto di responsabilità che l'ordinamento non consente e non potrebbe consentire senza tradire la propria funzione[94].

Se si guarda più da vicino a come, in concreto, l'uomo e il sistema si spartiscono il lavoro decisionale, ci si accorge che l'alternativa non è secca tra "decide la macchina" e "decide l'uomo". Esiste piuttosto una gradazione, che la riflessione sull'interazione uomo-macchina ha ormai messo a fuoco con una certa precisione. All'estremo di maggiore presidio umano si colloca il modello cosiddetto *human-in-the-loop*, in cui l'intervento della persona è parte integrante e necessaria del processo: nessuna decisione produce effetti finché un operatore non l'ha esaminata e confermata. In posizione intermedia si trova il modello *human-on-the-loop*, in cui il sistema opera in autonomia ma sotto la sorveglianza costante di un supervisore, che può intervenire per correggerlo o arrestarlo. All'estremo opposto sta il modello *human-out-of-the-loop*, in cui la macchina agisce senza alcun coinvolgimento umano diretto, ed è semmai l'uomo a doversi inserire ex post. Non sono distinzioni meramente descrittive: a ciascun grado corrisponde una diversa misura del controllo effettivo e, di riflesso, una diversa possibilità di attribuire la responsabilità di ciò che accade.

Collocare la funzione antiriciclaggio lungo questa scala aiuta a comprendere la scelta compiuta dal regolatore. Il modello verso cui spingono tanto la logica del sistema quanto le indicazioni più recenti è quello *human-in-the-loop*, o al più *human-on-the-loop* per le fasi puramente preparatorie: l'intelligenza artificiale può individuare, filtrare, ordinare

[93] LEMMA V., «Intelligenza Artificiale e sistemi di controllo: quali prospettive regolamentari?», cit. L'autore rileva che, mentre nei rapporti puramente artificiali viene in considerazione essenzialmente la rispondenza degli algoritmi ai principi di efficienza, nei rapporti misti si pongono questioni ulteriori, riferibili alle differenze e alle somiglianze dell'intelligenza artificiale rispetto ai ragionamenti umani.

[94] Sull'attribuzione dell'obbligo di segnalazione a soggetti determinati e sulle connesse responsabilità, cfr. D.Lgs. 21 novembre 2007, n. 231, artt. 35 e 58. Sull'impossibilità, allo stato del diritto, di configurare l'algoritmo quale autonomo centro di imputazione di responsabilità, cfr. LEMMA V., «Intelligenza Artificiale e sistemi di controllo: quali prospettive regolamentari?», cit.

per priorità, ma il passaggio dalla mera evidenza alla segnalazione deve rimanere nelle mani di una persona. Non è un caso che una parte della dottrina più attenta abbia insistito proprio sull'esigenza di "mettere l'uomo nel circuito" come risposta ai problemi sollevati dall'opacità dei sistemi, individuando nella presenza umana non un residuo del passato, ma un presidio necessario contro i rischi dell'automazione[95].

Su questo punto è intervenuto, da ultimo, il regolatore nazionale. Le Istruzioni UIF del dicembre 2025, come si è visto, ammettono l'impiego di sistemi di intelligenza artificiale nell'individuazione delle anomalie, a condizione che le anomalie da essi evidenziate siano controllate ed eventualmente validate con l'intervento umano[96]. La disposizione si presta a essere letta come la traduzione normativa della logica system-to-human: la macchina evidenzia, l'uomo valida. L'anomalia segnalata dall'algoritmo non è una conclusione, ma un'ipotesi di lavoro, che acquista rilevanza solo dopo essere passata attraverso il filtro del giudizio umano. Non è casuale che la stessa disposizione richiami, sullo sfondo, il regolamento europeo sull'intelligenza artificiale (AI Act), che al tema della sorveglianza umana sui sistemi ad alto rischio dedica una disciplina specifica, oggetto di analisi nel capitolo seguente.

Affermare che la decisione deve restare umana, tuttavia, non basta a garantire che il controllo sia reale. Qui si annida uno dei problemi più insidiosi dell'intero discorso, che la sola previsione normativa non è in grado di risolvere: il fenomeno dell'*automation bias*. Con questa espressione la letteratura sull'interazione uomo-macchina indica la tendenza dell'operatore a conformarsi acriticamente a ciò che il sistema gli suggerisce, ad assumere cioè come corretto l'output automatizzato senza sottoporlo a un vaglio autonomo. Il meccanismo ha una sua logica psicologica comprensibile: di fronte a un sistema che nella grande maggioranza dei casi si rivela affidabile, l'operatore sviluppa una fiducia che lo induce, quasi naturalmente, ad abbassare la guardia. E qui si cela il paradosso: quanto più il sistema è accurato, tanto più forte diventa la propensione a fidarsene, e tanto più raro l'esercizio di un controllo critico; sicché, nel momento eccezionale in cui la macchina sbaglia (un caso anomalo che il modello non aveva imparato a riconoscere) è anche il momento in cui l'operatore è meno predisposto ad accorgersene. La vigilanza umana rischia allora di rivelarsi più efficace sulla carta che nella pratica, trasformandosi da controllo sostanziale in ratifica formale di ciò che il sistema ha già stabilito.

Questo rischio si aggrava quando si considera la provenienza dei sistemi impiegati. I software di monitoraggio sono nella generalità dei casi progettati e programmati non all'interno della banca, ma da soggetti terzi (le *software-house*) estranei all'organizzazione dell'intermediario e ai controlli cui questo è sottoposto[97]. L'albero

[95] L'esigenza di "mettere l'uomo nel circuito" quale risposta ai problemi di opacità dei sistemi di intelligenza artificiale è sottolineata, nella letteratura richiamata dallo stesso LEMMA V., «Intelligenza Artificiale e sistemi di controllo: quali prospettive regolamentari?», cit., da ZETZSCHE D.A. – ARNER D.W. – BUCKLEY R.P. – TANG B., «Artificial Intelligence in Finance: Putting the Human in the Loop», in *University of Hong Kong Faculty of Law Research Paper*, 2020.

[96] UIF, *Istruzioni per la rilevazione e la segnalazione delle operazioni sospette*, Provvedimento del 18 dicembre 2025, Parte Prima, Sezione II.

[97] LEMMA V., «Intelligenza Artificiale e sistemi di controllo: quali prospettive regolamentari?», cit., dove si osserva che le azioni indirette dell'intelligenza naturale sono il risultato di scelte predeterminate da un

decisionale che governa il comportamento del sistema, cioè la logica in base alla quale esso distingue il sospetto dal lecito, è dunque frutto di scelte compiute da chi il software lo produce, non da chi lo utilizza. Da qui un rischio che la dottrina definisce di eterodirezione: l'intermediario vigilato rischia di trovarsi, di fatto, guidato nelle proprie valutazioni da criteri elaborati altrove, da soggetti che sfuggono al perimetro della vigilanza e che non rispondono degli esiti delle segnalazioni[98].

Il rilievo è prezioso, perché mostra come la relazione *system-to-human* non si esaurisca nel rapporto tra l'analista e lo schermo che ha davanti, ma si estenda a monte, fino a chi quel sistema lo ha concepito. La validazione umana richiesta dalle Istruzioni UIF, per essere effettiva e non una formalità, presuppone che l'operatore sia in grado di comprendere e mettere in discussione ciò che il sistema gli propone. Ma questa possibilità si assottiglia proprio quando il sistema è più opaco e la sua logica è stata definita da terzi: se l'analista non è messo in condizione di capire perché l'algoritmo ha segnalato una certa operazione, la sua validazione scivola verso la ratifica, e *l'automation bias* trova terreno fertile. Le due criticità — l'opacità del modello e la sua origine esterna — si saldano così in un unico nodo, che tocca insieme il piano dell'organizzazione interna e quello, più ampio, della vigilanza.

Da un lato, l'ordinamento esige che la decisione resti umana e che il sistema sia sottoposto a validazione; dall'altro, l'efficacia di questa validazione dipende da una condizione — la comprensibilità del funzionamento del sistema — che i modelli più avanzati tendono a negare, e che si complica ulteriormente quando la logica del sistema è stata scritta da un fornitore esterno. Il principio della centralità del giudizio umano rischia allora di rimanere un'affermazione di facciata se non è accompagnato dagli strumenti che lo rendono concretamente esercitabile: la capacità di comprendere la decisione della macchina e la possibilità di individuare chi risponde quando quella decisione si rivela errata.

2.4 Verso il problema della "black box": efficacia tecnica e opacità decisionale

Il percorso compiuto in questo capitolo ha mostrato una traiettoria abbastanza netta. Si è partiti dai limiti dei sistemi tradizionali — la loro rigidità, l'incapacità di apprendere, la produzione massiccia di falsi positivi — per illustrare come l'intelligenza artificiale prometta di superarli, affinando la capacità di individuazione, riducendo il rumore di fondo e restituendo all'operatore umano un ruolo di analisi qualificata anziché di mero smistamento. Fin qui il bilancio parrebbe schiacciante a favore della nuova tecnologia, se non fosse per una tensione emersa a più riprese: quanto più i modelli diventano potenti, tanto meno risultano trasparenti. L'albero decisionale, semplice e leggibile, cede il passo

algoritmo ideato da persone fisiche estranee all'organizzazione dell'intermediario e impiegate presso le cosiddette *software-house*.

[98] LEMMA V., «Intelligenza Artificiale e sistemi di controllo: quali prospettive regolamentari?», cit., in ordine al rischio di eterodirezione e ai limiti della disciplina dell'*outsourcing* rispetto ai fornitori di sistemi di intelligenza artificiale, la cui attività incide sull'albero decisionale posto a fondamento delle valutazioni dell'intermediario.

a foreste, sistemi di *gradient boosting* e reti neurali via via più complessi e più difficili da interpretare, fino alle reti neurali su grafo, che colgono schemi relazionali sfuggenti a qualunque regola ma sono tra le architetture meno leggibili.

Questa relazione fra complessità e intelligibilità va però precisata, perché viene spesso presentata come una legge inderogabile mentre è, più esattamente, una tendenza. Nei modelli più complessi l'opacità è effettivamente inscindibile dal modo in cui sono costruiti: la conoscenza appresa vi è distribuita su una moltitudine di parametri privi di significato autonomo, e nessun accorgimento successivo può restituirle la forma di un ragionamento. Nulla obbliga, però, a ricorrere a modelli di quel tipo. Come si è visto, sui dati tabellari dell'operatività bancaria gli approcci basati su alberi conservano prestazioni competitive, e perfino nell'analisi delle reti i modelli più leggibili reggono il confronto con quelli opachi. L'opacità è dunque il prezzo di una determinata scelta progettuale, non una condizione imposta dalla materia. Quando quella scelta viene compiuta, si ottiene ciò che va sotto il nome di *black box* ("scatola nera"): un sistema che fornisce un risultato senza esplicitare, in termini comprensibili a un essere umano, il percorso logico che a quel risultato ha condotto[99].

Nella maggior parte dei contesti applicativi, questa opacità è un problema tollerabile: a chi utilizza un servizio di raccomandazione musicale importa poco di sapere perché il sistema gli abbia suggerito un certo brano. Nel campo antiriciclaggio, invece, l'opacità cessa di essere un dettaglio tecnico e diventa un problema giuridico di primo piano. Le ragioni sono state anticipate nel corso del capitolo, ma conviene raccoglierle, perché è su di esse che si costruisce il capitolo seguente. La segnalazione di un'operazione sospetta non è un atto neutro: incide sulla posizione del cliente, mette in moto l'apparato di *intelligence* finanziaria e può costituire il primo anello di una catena che conduce fino al procedimento penale. Un atto di tale rilievo esige di essere motivato. Come si è visto sin dal primo capitolo, le stesse Istruzioni UIF impongono al delegato di esporre con chiarezza i motivi del sospetto, ricostruendo il percorso logico che ha condotto alla segnalazione. Inoltre, un'analoga esigenza di giustificazione richiede che la decisione sia archiviata in modo tracciabile e ricostruibile a posteriori[100].

Qui si consuma il cortocircuito. Se la segnalazione trae origine, in tutto o in parte, dall'evidenza prodotta da un sistema che non è in grado di spiegare le proprie conclusioni, come può il delegato motivare adeguatamente la propria decisione? Come può sostenere, davanti all'autorità di vigilanza o in una sede giudiziaria, le ragioni di un sospetto la cui scintilla iniziale è scoccata all'interno di una scatola che nessuno sa aprire? E, sul versante opposto, come può giustificare l'archiviazione di un caso che il sistema aveva segnalato, se non è in grado di comprendere perché il sistema lo avesse segnalato? Il requisito della motivazione, che l'ordinamento antiriciclaggio pone a presidio della qualità e della

[99] Sul fenomeno della *black box* quale opacità del percorso decisionale dei sistemi di intelligenza artificiale più complessi, e sulla correlazione inversa tra accuratezza dei modelli e loro interpretabilità, cfr. KUTE D.V. – PRADHAN B. – SHUKLA N. – ALAMRI A., «Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering – A Critical Review», in *IEEE Access*, 2021.

[100] UIF – Unità di Informazione Finanziaria per l'Italia, *Istruzioni per la rilevazione e la segnalazione delle operazioni sospette*, Provvedimento del 18 dicembre 2025, Parte Prima, Sezione IV (Segnalazione delle operazioni sospette) e Sezione III (Esame delle anomalie), in ordine all'obbligo di rappresentare i motivi del sospetto e di conservare traccia delle valutazioni svolte; disponibile su: uif.bancaditalia.it

controllabilità delle decisioni, entra in rotta di collisione con l'opacità dei modelli più sofisticati[101].

Si comprende, allora, perché la centralità del giudizio umano affermata nel paragrafo precedente non basti, da sola, a sciogliere il nodo. Non è sufficiente stabilire che sia l'uomo a decidere, se poi l'uomo decide sulla base di un'evidenza che non è in grado di comprendere fino in fondo. Il controllo umano, per essere effettivo, presuppone la comprensibilità di ciò che viene controllato; e la comprensibilità, a sua volta, chiama in causa il tema della spiegabilità dei sistemi di intelligenza artificiale: la capacità, cioè, di rendere intelligibile il percorso che porta l'algoritmo a una determinata conclusione. È attorno a questo tema, e alle risposte che l'ordinamento europeo e nazionale iniziano a fornire, che ruota il terzo e ultimo capitolo. Vi si affronteranno il requisito della spiegabilità, come condizione di legittimità dell'impiego dell'intelligenza artificiale nel processo di segnalazione; la questione della responsabilità dell'intermediario e degli altri soggetti coinvolti; e il quadro regolatorio disegnato dall'AI Act.

[101] Sulla tensione tra il requisito di motivazione degli atti e l'opacità dei modelli di intelligenza artificiale nel settore antiriciclaggio, cfr. KUTE D.V. e altri, «Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering», cit., ove si evidenzia la necessità di tecniche di intelligenza artificiale spiegabile (*explainable AI*) per garantire trasparenza e responsabilità nei sistemi di rilevazione del riciclaggio.

Capitolo 3 - Profili di governance e il problema della “Black Box”

Il capitolo precedente si è chiuso su una tensione irrisolta. Da un lato l’intelligenza artificiale offre risposte concrete alle inefficienze del monitoraggio tradizionale; dall’altro i modelli più sofisticati sono anche quelli che meno sanno rendere conto di sé. Finché il discorso resta sul terreno dell’efficienza, questa opacità può apparire un prezzo accettabile: se il sistema individua meglio le anomalie, poco importa che il suo ragionamento sia difficile da ricostruire. Ma nel campo antiriciclaggio il discorso non resta su quel terreno. La segnalazione di un’operazione sospetta è un atto che produce effetti giuridici, incide sulla posizione di una persona e può costituire il primo passo di un percorso che conduce all’indagine penale. Un atto simile non può fondarsi su ragioni che nessuno è in grado di esporre.

È qui che il problema tecnico diventa problema di governance. Dire che la decisione deve restare umana, come si è argomentato, non basta: occorre chiedersi a quali condizioni quella decisione sia effettivamente umana e non una ratifica di ciò che il sistema ha già stabilito. Occorre stabilire chi risponda quando la macchina sbaglia, e se possa rispondere soltanto chi l’ha utilizzata o anche chi l’ha costruita. Occorre infine verificare se l’ordinamento giuridico abbia predisposto strumenti adeguati, o se permanga quel vuoto che la dottrina denunciava già prima che il legislatore europeo intervenisse.

Il capitolo affronta questi interrogativi in tre passaggi. Il primo muove dal requisito della motivazione, per verificare cosa significhi, sul piano tecnico e su quello giuridico, pretendere che una decisione algoritmica sia spiegabile, e quanto le tecniche oggi disponibili siano all’altezza di tale pretesa. Il secondo si sposta sul terreno della responsabilità, interrogandosi su come si distribuiscano gli obblighi e le conseguenze del loro inadempimento tra l’intermediario che impiega il sistema, i soggetti che vi operano e i fornitori che quel sistema lo hanno progettato. Il terzo ricostruisce il quadro regolatorio emergente, con particolare attenzione al regolamento europeo sull’intelligenza artificiale e al modo in cui esso si innesta sulla disciplina antiriciclaggio e sugli orientamenti delle autorità nazionali.

Il filo che li tiene insieme è quello che percorre l’intera tesi. Se l’intelligenza artificiale deve restare uno strumento di supporto e non un sostituto del giudizio, allora la questione decisiva non è quanto essa sia accurata, ma quanto sia governabile: se cioè l’ordinamento riesca a mantenere, su un processo sempre più automatizzato, quel controllo umano che della responsabilità è il presupposto.

3.1 La necessità di motivare la segnalazione: il requisito della “spiegabilità” delle decisioni algoritmiche

Il termine spiegabilità viene usato, nel dibattito corrente, con almeno due significati da tenere distinti, perché rispondono a domande diverse e hanno diversi interlocutori.

Vi è anzitutto una spiegabilità in senso giuridico, che attiene all'obbligo di motivare un atto: chi assume una decisione deve poterne esporre le ragioni a chi ha titolo per chiederle. È un requisito antico, che nell'antiriciclaggio si traduce nell'onere, gravante sul delegato SOS, di rappresentare i motivi del sospetto in modo che l'autorità possa comprenderli e verificarli. Questa esigenza esiste indipendentemente dalla tecnologia impiegata: valeva quando le segnalazioni nascevano dall'intuizione di un operatore di sportello e vale oggi.

Poi vi è una spiegabilità in senso tecnico, che riguarda la possibilità di ricostruire il funzionamento di un modello di apprendimento automatico e di comprendere quali elementi lo abbiano condotto a un determinato risultato.

Le due nozioni non coincidono, ma nell'ambito che qui interessa finiscono per intrecciarsi strettamente: quando il percorso che porta alla segnalazione passa attraverso un algoritmo, l'adempimento dell'obbligo giuridico di motivazione dipende, almeno in parte, dalla disponibilità di strumenti tecnici che rendano intelligibile l'output del sistema. Il paragrafo procede perciò su entrambi i piani, muovendo dal fondamento normativo dell'obbligo di motivare per poi verificare in che misura le tecniche di intelligenza artificiale spiegabile siano in grado di sostenerlo.

3.1.1 L'obbligo di motivare il sospetto: il fondamento normativo

Prima di interrogarsi su quanto un algoritmo sia in grado di spiegare le proprie conclusioni, si deve stabilire perché una spiegazione sia richiesta. Il punto non è scontato: in molti ambiti applicativi l'opacità di un sistema automatizzato è un fastidio tecnico, non un problema giuridico. Nel campo delle segnalazioni di operazioni sospette, invece, l'esigenza di rendere conto del ragionamento non è un'aspirazione di buona pratica, ma un obbligo che deriva direttamente dalla disciplina. È da qui che occorre partire.

Già la norma primaria contiene, in filigrana, un requisito di giustificazione. L'articolo 35 del decreto antiriciclaggio non impone di segnalare ogni operazione insolita, bensì quelle rispetto alle quali il soggetto obbligato sappia, sospetti o abbia motivi ragionevoli per sospettare che sia in corso un'operazione di riciclaggio o di finanziamento del terrorismo[102]. Un sospetto ragionevole è, per definizione, un sospetto che poggia su ragioni; e ragioni che devono poter essere esposte, perché altrimenti la ragionevolezza resterebbe un'affermazione priva di verifica. La norma, in altri termini, non si accontenta che il segnalante abbia avuto un'intuizione corretta: pretende che quell'intuizione sia argomentabile.

Ciò che nella fonte primaria è implicito diventa esplicito e dettagliato nelle Istruzioni della UIF. Il provvedimento di dicembre 2025 stabilisce che i destinatari, all'esito del processo valutativo, selezionino le operazioni sospette e le segnalino esplicitando gli elementi informativi rilevanti a sostegno delle proprie valutazioni, così da rappresentare compiutamente le circostanze soggettive e oggettive su cui il sospetto si fonda[103]. La segnalazione, dunque, non è un atto di mera trasmissione di dati: è un atto argomentativo,

[102] D.Lgs. 21 novembre 2007, n. 231, art. 35, comma 1.
[103] UIF – Unità di Informazione Finanziaria per l'Italia, *Istruzioni per la rilevazione e la segnalazione delle operazioni sospette*, Provvedimento del 18 dicembre 2025, Parte Prima, Sezione IV, disponibile su: uif.bancaditalia.it.

che deve mettere l'autorità nelle condizioni di comprendere non soltanto cosa sia accaduto, ma perché l'intermediario lo abbia ritenuto anomalo.

La struttura stessa del modello segnaletico riflette questa impostazione. Accanto ai dati strutturati (soggetti, operatività, rapporti e legami) le Istruzioni prevedono una parte descrittiva in forma libera, articolata in due sezioni distinte e non sovrapponibili: la descrizione dell'operatività sospetta e i motivi del sospetto[104]. La prima ha funzione ricostruttiva e serve a rendere intelligibili i flussi finanziari; la seconda, che qui più interessa, ha funzione propriamente motivazionale. In essa il destinatario è chiamato a descrivere le ragioni sottostanti alla segnalazione, ovvero il processo logico-deduttivo seguito con riguardo all'individuazione delle anomalie, alla loro valutazione e alle circostanze che lo hanno indotto a ritenere sussistente il sospetto[105]. Non è richiesto un esito, ma un percorso. All'intermediario si domanda di esporre come sia arrivato dove è arrivato.

Le Istruzioni si spingono oltre, precisando cosa questa motivazione non possa essere. È espressamente stabilito che i motivi del sospetto non debbano risolversi in un mero rinvio agli indicatori di anomalia, agli schemi o alle comunicazioni della UIF, né costituirne la replica: tali strumenti possono agevolare l'elaborazione della motivazione, purché essa contestualizzi e circoscriva gli elementi di anomalia rispetto all'operatività concretamente osservata e al profilo del soggetto cui si riferisce[106]. Il legislatore secondario, in sostanza, vieta la motivazione per formula: non basta invocare la categoria astratta sotto cui l'anomalia ricade, occorre spiegare perché in quel caso specifico, per quel cliente, quella condotta risulti ingiustificabile. È una prescrizione che va letta insieme all'avvertenza, contenuta nello stesso provvedimento, di evitare meri approcci compilativi[107].

Vale la pena osservare che l'onere di giustificazione non grava soltanto sulla decisione di segnalare. Anche la scelta opposta di archiviare, ritenendo il sospetto insussistente, deve poter essere ricostruita. Le Istruzioni riconoscono che il processo valutativo possa legittimamente concludersi con l'esclusione del sospetto, persino in presenza di fattispecie astrattamente riconducibili a indicatori di anomalia; ma richiedono che il destinatario adotti accorgimenti idonei ad agevolare la ricostruzione a posteriori delle valutazioni effettuate, conservando una traccia, anche sintetica, delle ragioni ritenute sufficienti a escluderlo[108]. L'obbligo di motivazione è quindi bidirezionale: copre tanto ciò che si segnala quanto ciò che si lascia cadere. Ed è comprensibile che sia così, se si considera che l'omessa segnalazione è presidiata da sanzione e che, in sede ispettiva,

[104] UIF, *Istruzioni* del 18 dicembre 2025, cit., Parte Terza, Sezione II, punto 2, lett. C. Il provvedimento precisa che non sono ammessi contenuti identici nelle due parti e che non è consentito omettere la compilazione di alcuna di esse.
[105] UIF, *Istruzioni* del 18 dicembre 2025, cit., Parte Terza, Sezione II, punto 2, lett. C.
[106] UIF, *Istruzioni* del 18 dicembre 2025, cit., Parte Terza, Sezione II, punto 2, lett. C.
[107] UIF, *Istruzioni* del 18 dicembre 2025, cit., Parte Prima, Sezione IV, ove si raccomanda di evitare approcci meramente compilativi, quali l'elencazione indiscriminata di operazioni non collegate all'operatività ritenuta sospetta.
[108] UIF, *Istruzioni* del 18 dicembre 2025, cit., Parte Prima, Sezione III.

l'assenza di una motivazione scritta a fronte di un'operazione oggettivamente anomala espone l'intermediario a conseguenze rilevanti[109].

Da questo insieme di previsioni emerge una fisionomia precisa del sistema. La segnalazione di operazione sospetta è concepita come un atto motivato, la cui qualità si misura non sulla mole di dati trasmessi ma sulla tenuta del ragionamento che li accompagna. Non si tratta di un formalismo: la motivazione svolge funzioni concrete e diverse tra loro. Verso l'autorità, consente all'analista della UIF di inquadrare rapidamente il fenomeno e di orientare l'approfondimento, evitando che il valore informativo della segnalazione si disperda. Verso l'intermediario, costituisce la prova della diligenza impiegata e il presidio della sua posizione in caso di controllo. Verso il soggetto segnalato, infine, rappresenta una garanzia indiretta, poiché impone che l'attenzione delle autorità si appunti su di lui in forza di ragioni esplicitate e non di automatismi.

È a questo punto che il quadro entra in tensione con quanto si è visto nel capitolo precedente. Se la motivazione deve consistere nell'esposizione di un percorso logico-deduttivo, e se quel percorso ha avuto origine nell'evidenza prodotta da un sistema che non è in grado di esplicitare le proprie ragioni, il delegato si trova in una posizione scomoda. Può descrivere ciò che ha verificato dopo *l'alert*, ricostruendo autonomamente il quadro; ma non può spiegare perché il sistema abbia segnalato quel cliente e non un altro, se il sistema stesso non glielo dice. Il rischio, concreto, è che la motivazione si riduca a una razionalizzazione a posteriori: un ragionamento costruito per giustificare una conclusione già raggiunta altrove, che ha la forma della motivazione ma non la sostanza. Ed è appena il caso di notare che una motivazione così confezionata violerebbe proprio il divieto, sopra richiamato, di ricondurre il sospetto a formule generiche anziché al caso concreto. Il problema della spiegabilità algoritmica, prima ancora di essere una questione tecnica, si presenta dunque come un problema di tenuta dell'obbligo motivazionale.

3.1.2 La spiegabilità tecnica: cosa può e cosa non può offrire l'intelligenza artificiale

Stabilito che l'ordinamento pretende una motivazione, resta da capire se la tecnologia sia in grado di fornirne il materiale. La domanda va posta correttamente, perché su questo punto circola un equivoco diffuso: si tende a immaginare che il sistema "sappia" perché ha segnalato un cliente e semplicemente non lo comunichi, come se si trattasse di un problema di interfaccia, di un difetto di traduzione fra il linguaggio della macchina e quello dell'uomo. Le cose stanno diversamente, e la differenza è sostanziale.

Nei modelli più complessi la conoscenza acquisita durante l'addestramento non è depositata in regole, ma distribuita su un numero enorme di parametri numerici, ciascuno dei quali, isolatamente considerato, è privo di significato. Quando il sistema attribuisce a un'operazione una data probabilità di essere sospetta, non ha percorso una catena di inferenze traducibile in parole: ha eseguito un calcolo che combina quei parametri secondo una geometria che nessuno ha progettato esplicitamente, e che è emersa

[109] D.Lgs. 21 novembre 2007, n. 231, art. 58, in materia di sanzioni per omessa segnalazione di operazioni sospette.

dall'esposizione ai dati. Non esiste, dentro il modello, una motivazione che qualcuno stia trattenendo: esiste un risultato, e null'altro. Di conseguenza: neppure chi il sistema lo ha progettato è in grado, di norma, di spiegare una singola decisione. Il programmatore conosce l'architettura scelta, i dati impiegati per l'addestramento e i criteri di ottimizzazione, ma non può dire perché quel modello abbia segnalato quel cliente in quel giorno. L'opacità non è dunque un difetto di comunicazione fra la macchina e l'uomo: è una caratteristica strutturale del modo in cui il modello è costruito[110].

Occorre però evitare la semplificazione opposta, secondo cui ogni sistema di apprendimento automatico sarebbe una scatola nera. Come si è visto illustrando i modelli, l'opacità cresce con la complessità: un albero decisionale, pur apprendendo dai dati, resta perfettamente leggibile, perché il percorso che conduce a una foglia è ricostruibile passo per passo e traducibile in una sequenza di condizioni. È salendo verso le foreste, i sistemi di *gradient boosting* e soprattutto le reti neurali profonde che la leggibilità si perde progressivamente. L'opacità non è il prezzo dell'apprendimento in sé, ma della sofisticazione del modello impiegato. La precisazione non è oziosa, perché ne discende un punto che tornerà nel prosieguo: se un sistema è opaco, lo è in conseguenza di una scelta, non di una fatalità.

Alla difficoltà si è cercato di rispondere sviluppando un insieme di tecniche che vanno sotto il nome di intelligenza artificiale spiegabile (*explainable artificial intelligence*, comunemente abbreviata in XAI). Conviene innanzitutto introdurre una distinzione che la letteratura tecnica considera fondamentale e che ha riflessi diretti sul piano giuridico: quella tra interpretabilità intrinseca e spiegabilità *post hoc*. Nel primo caso il modello è leggibile per costruzione, perché la sua struttura stessa rende visibile il percorso decisionale: è il caso dell'albero, o di una semplice regressione. Nel secondo, il modello resta opaco e si affianca a esso un meccanismo separato, che cerca di ricostruirne il comportamento dall'esterno osservando come l'output vari al variare degli ingressi. La quasi totalità degli strumenti oggi impiegati nella pratica appartiene alla seconda categoria, per una ragione facilmente intuibile: consente di conservare i modelli più performanti aggiungendovi uno strato esplicativo, senza rinunciare all'accuratezza.

Due sono gli strumenti divenuti standard in questo campo. Il primo, noto con la sigla LIME, muove da un'intuizione elegante: per spiegare una singola decisione non occorre comprendere il modello nella sua globalità, ma soltanto nell'intorno ristretto del caso esaminato. Il metodo genera dunque una serie di casi artificiali simili a quello in esame, osserva come il sistema li classifichi e costruisce su tali osservazioni un modello semplificato e leggibile, tipicamente lineare, che approssima il comportamento del sistema complesso in quella zona circoscritta, restituendo così un'indicazione dei fattori che hanno pesato di più su quel caso[111]. Il secondo, denominato SHAP, muove invece

[110] Sul carattere strutturale dell'opacità dei modelli complessi e sull'impossibilità di ricondurne l'esito a una catena di inferenze intelligibile, cfr. KUTE D.V. – PRADHAN B. – SHUKLA N. – ALAMRI A., «Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering – A Critical Review», in *IEEE Access*, 2021.

[111] RIBEIRO M.T. – SINGH S. – GUESTRIN C., «"Why Should I Trust You?": Explaining the Predictions of Any Classifier», in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, dove è presentata la metodologia *Local Interpretable Model-agnostic Explanations* (LIME).

dalla teoria dei giochi cooperativi e affronta il problema come una questione di attribuzione: dato un risultato prodotto dal concorso di più variabili, quanto ha contribuito ciascuna? La risposta è costruita distribuendo il risultato fra le variabili di ingresso secondo il criterio di equità formalizzato dal valore di Shapley, che assegna a ciascun fattore un peso proporzionale al contributo marginale che apporta in tutte le possibili combinazioni[112].

In entrambi i casi l'esito è una graduatoria dei fattori che avrebbero determinato la decisione. Trasposto nel nostro ambito, il delegato SOS che apra il fascicolo di un alert non troverebbe più soltanto l'indicazione che il cliente è stato segnalato, ma anche l'indicazione che a pesare sono stati, per esempio: la frequenza dei versamenti in contante nell'ultimo trimestre, la loro concentrazione in giornate contigue e la sproporzione rispetto ai redditi noti. È un'informazione preziosa che consente di orientare l'istruttoria.

Il progresso è reale e non va sottovalutato: disporre di una simile graduatoria è incomparabilmente meglio che ricevere un allarme muto. Occorre però comprendere con esattezza che cosa questi strumenti offrano, perché il rischio di sopravvalutarli è concreto e ha ricadute proprio sul terreno che qui interessa.

Il punto essenziale è che tali tecniche non aprono la scatola: ne osservano il comportamento dall'esterno e ne costruiscono un modello semplificato. Ciò che restituiscono, dunque, non è il ragionamento del sistema — che, come si è visto, in senso proprio non esiste — ma un'approssimazione plausibile del suo comportamento, elaborata da un meccanismo diverso, con una logica diversa e su una base informativa parziale, costituita dalle sole risposte del modello a un campione di ingressi[113]. Fra la spiegazione e ciò che essa spiega permane uno scarto incolmabile, che non dipende dalla perfettibilità degli strumenti ma dalla loro natura di surrogati.

Da questa natura discendono limiti documentati da richiamare, perché ciascuno di essi ha un corrispettivo sul piano dell'affidabilità giuridica.

Vi è anzitutto un problema di stabilità. Poiché metodi come LIME si fondano su perturbazioni casuali degli ingressi, la spiegazione fornita può variare fra esecuzioni successive sullo stesso caso; e piccole differenze nell'input possono produrre spiegazioni sensibilmente diverse anche quando la previsione resta identica[114]. In un contesto nel quale la coerenza del ragionamento è requisito di legittimità, una spiegazione che muti

---

[112] LUNDBERG S.M. – LEE S.I., «A Unified Approach to Interpreting Model Predictions», in *Advances in Neural Information Processing Systems*, 2017, ove è presentata la metodologia *SHapley Additive exPlanations* (SHAP).

[113] Sulla natura surrogata delle spiegazioni *post hoc*, che non possono essere pienamente fedeli al modello che pretendono di spiegare, cfr. RUDIN C., «Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead», in *Nature Machine Intelligence*, vol. 1, 2019.

[114] Sull'instabilità delle spiegazioni di LIME fra esecuzioni ripetute sullo stesso caso, cfr. VISANI G. – BAGLI E. – CHESANI F. – POLUZZI A. – CAPUZZO D., «Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models», in *Journal of the Operational Research Society*, vol. 73, n. 1, 2022; sulla sensibilità delle spiegazioni a piccole variazioni dell'input, cfr. ALVAREZ-MELIS D. – JAAKKOLA T.S., «On the Robustness of Interpretability Methods», 2018 ICML Workshop on Human Interpretability in Machine Learning (WHI 2018), Stoccolma, 2018.

arbitrariamente fra casi analoghi è problematica: non soltanto perché inaffidabile, ma perché crea l'impressione di una trasparenza che in realtà non fornisce.

Vi è poi un problema di fedeltà. L'approssimazione locale è accurata soltanto dove il comportamento del modello è sufficientemente regolare, e tende a deteriorarsi in prossimità di confini decisionali complessi. Spesso i casi più interessanti dal punto di vista investigativo sono quelli in cui l'operatività si colloca al limite fra il lecito e il sospetto. Proprio là dove la spiegazione servirebbe di più, essa è meno attendibile.

Vi è infine un problema di manipolabilità, dimostrato in modo netto dalla ricerca. Uno studio ormai classico ha provato che i metodi fondati su perturbazioni, come LIME e SHAP, non sono affidabili, poiché è possibile costruire sistemi che occultano deliberatamente i propri criteri, inducendo tali metodi a produrre spiegazioni innocue che non riflettono le distorsioni sottostanti[115]. La circostanza è rilevante sotto due profili distinti: mostra che una spiegazione può essere tecnicamente ineccepibile e sostanzialmente falsa, e apre l'interrogativo su chi risponda quando il sistema è progettato da un terzo che quelle spiegazioni le confeziona.

Su queste basi si è formata in dottrina informatica una posizione più radicale, che merita di essere riferita perché ha implicazioni dirette sul discorso giuridico. Essa sostiene che, nei contesti nei quali una decisione incide significativamente sulle persone, non ci si dovrebbe affidare a spiegazioni costruite a posteriori di modelli opachi, bensì impiegare modelli intrinsecamente interpretabili, la cui logica sia leggibile per costruzione; e che la presunta necessità di sacrificare l'interpretabilità in nome dell'accuratezza sia, in molti ambiti applicativi, assai meno fondata di quanto si creda, trattandosi spesso di un assunto non verificato[116].

L'argomento sposta la questione dal piano del rimedio a quello della prevenzione: se il problema è che la scatola non si apre, una via è non costruirla chiusa. E ha un rilievo particolare nel nostro campo, per due ragioni. La prima è che, come si è visto nel capitolo precedente, sui dati tabellari tipici dell'operatività bancaria i modelli ad alberi conservano prestazioni competitive rispetto alle architetture più opache: il sacrificio di accuratezza che si dovrebbe accettare per guadagnare trasparenza è, in questo ambito, verosimilmente contenuto. La seconda è che la scelta del modello cessa così di essere una decisione meramente tecnica, rimessa alla funzione informatica o al fornitore, e diventa una scelta di *governance*, che incide sulla capacità dell'intermediario di adempiere ai propri obblighi. È un punto che sarà ripreso trattando della responsabilità nel prossimo paragrafo.

Un ulteriore profilo, spesso trascurato, riguarda il destinatario della spiegazione. Le tecniche descritte sono state concepite in ambito informatico, prevalentemente per finalità di diagnosi: consentire a chi sviluppa un modello di verificarne il comportamento, individuare distorsioni, correggere errori. Il loro destinatario naturale è dunque il tecnico.

---

[115] SLACK D. – HILGARD S. – JIA E. – SINGH S. – LAKKARAJU H., «Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods», in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, 2020.

[116] RUDIN C., «Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead», cit.

Ma nel nostro contesto le spiegazioni servono a soggetti diversi e per scopi diversi, e questa divaricazione non è priva di conseguenze.

È anzitutto utile all'analista, che deve decidere se approfondire una posizione: a lui occorre un'indicazione operativa su dove concentrare la verifica. Serve poi al delegato SOS, che deve costruire la motivazione della segnalazione: a lui occorre materiale argomentativo spendibile, formulato in termini comprensibili a chi non conosce il funzionamento del modello. È necessario inoltre alla funzione di controllo e all'autorità di vigilanza, che devono valutare se il processo sia stato governato correttamente: a loro occorre poter ricostruire non tanto la singola decisione, quanto la logica complessiva del sistema e la sua coerenza con l'approccio basato sul rischio. Serve infine, potenzialmente, in sede giudiziaria, dove il grado di robustezza richiesto è ancora superiore.

Che un unico strumento, nato per finalità diagnostiche, possa soddisfare simultaneamente esigenze tanto diverse è tutt'altro che scontato. Una graduatoria di fattori può essere sufficiente per orientare un approfondimento istruttorio e del tutto inadeguata a sostenere un contraddittorio.

Si può ora tornare alla domanda da cui il paragrafo ha preso le mosse. Le tecniche descritte sono sufficienti a soddisfare l'obbligo di rappresentare i motivi del sospetto? La risposta non può essere netta, ma alcune considerazioni si impongono.

Una graduatoria dei fattori rilevanti offre al delegato un punto di partenza prezioso: gli indica dove guardare e gli consente di verificare, sui dati di cui dispone, se quei fattori reggano effettivamente. In questo senso la spiegazione tecnica non sostituisce la motivazione giuridica, ma la alimenta, fornendo materiale che l'analista deve poi sottoporre a un vaglio proprio. E il vaglio è indispensabile per una ragione che le stesse fonti tecniche segnalano con chiarezza: una spiegazione riferisce ciò che il modello ha appreso, non ciò che è vero. Se il sistema ha assimilato una correlazione priva di fondamento sostanziale, perché presente nei dati storici per ragioni accidentali, la spiegazione la restituirà fedelmente, presentandola come un fattore rilevante. Stabilire se quella correlazione abbia un senso economico e giuridico è compito che nessuno strumento automatico può assolvere, e che richiede la conoscenza del cliente, del settore e del contesto.

Vi è poi una differenza di natura fra i due piani. La spiegazione tecnica è descrittiva: dice quali fattori hanno pesato. La motivazione giuridica è invece valutativa: deve dire perché quei fattori, in quel caso, giustifichino un sospetto di riciclaggio. Il passaggio dall'una all'altra non è automatico, ed è precisamente il luogo in cui si esercita il giudizio professionale. Affermare che un conto è stato segnalato perché presentava versamenti frequenti e sproporzionati non è ancora una motivazione: lo diventa quando si spiega perché quella sproporzione, alla luce di quanto si conosce del cliente e delle giustificazioni da lui addotte, non trovi spiegazione lecita. Le stesse Istruzioni della UIF, del resto, esigono che la motivazione contestualizzi e circoscriva gli elementi di anomalia rispetto all'operatività osservata e al profilo del soggetto: una graduatoria di fattori, per sua natura, non contestualizza nulla.

Ne deriva che la spiegabilità tecnica, per come oggi si presenta, è una condizione utile ma non sufficiente. Utile, perché senza di essa il delegato SOS non avrebbe alcun appiglio e la sua motivazione sarebbe inevitabilmente una costruzione a posteriori scollegata dall'evidenza che ha innescato il processo. Non sufficiente, perché una spiegazione instabile, approssimata e potenzialmente infedele non può, da sola, sorreggere un atto destinato a produrre conseguenze giuridiche.

Resta quindi confermato, e anzi rafforzato, quanto si è argomentato nel capitolo precedente: la validazione umana non è una formalità aggiungibile a piacere, ma il momento in cui l'indizio statistico prodotto dalla macchina viene tradotto in ragione giuridicamente spendibile. Con due avvertenze, che i paragrafi successivi svilupperanno. La prima è che perché quella traduzione sia possibile occorre che chi la compie disponga degli strumenti e delle competenze per compierla: un delegato SOS posto di fronte a una graduatoria di fattori che non è in grado di interpretare non valida alcunché. La seconda è che, quando la traduzione fallisce, quando cioè viene segnalato chi non doveva esserlo, o non viene segnalato chi doveva, occorre sapere chi ne risponda: se soltanto chi ha compiuto la valutazione finale, o anche chi ha costruito il sistema che l'ha orientata.

3.2 La responsabilità dell'intermediario e del delegato SOS

3.2.1 Il quadro sanzionatorio dell'antiriciclaggio e la sua struttura

Si parte da come l'ordinamento antiriciclaggio distribuisce oggi la responsabilità, perché è su quel telaio che le nuove tecnologie si innestano.

Il decreto antiriciclaggio individua i destinatari degli obblighi all'articolo 3 e presidia l'inadempimento con un apparato sanzionatorio articolato. Per quanto qui rileva, l'omessa segnalazione di operazioni sospette è sanzionata dall'articolo 58, che colpisce i soggetti obbligati che, in violazione dell'articolo 35, omettano di effettuare la segnalazione, con un aggravamento delle conseguenze nei casi di violazioni gravi, ripetute o sistematiche[117]. Accanto alla responsabilità dell'ente si colloca quella delle persone fisiche che esercitano funzioni di amministrazione, direzione e controllo, secondo un modello che coinvolge tanto la struttura quanto chi la governa. Per gli intermediari vigilati opera inoltre il regime specifico dell'articolo 62, che per le violazioni gravi, ripetute o sistematiche prevede sanzioni amministrative di importo molto più elevato[118].

Due caratteristiche di questo impianto meritano di essere sottolineate perché condizionano tutto il ragionamento che segue.

La prima è che la responsabilità è ancorata a un obbligo di comportamento, non di risultato. Non si risponde per non aver individuato un riciclaggio (esito che dipende da

[117] D.Lgs. 21 novembre 2007, n. 231, art. 58, in materia di sanzioni per l'omessa segnalazione di operazioni sospette.
[118] D.Lgs. 21 novembre 2007, n. 231, art. 62, recante disposizioni sanzionatorie specifiche per i soggetti obbligati vigilati, applicabili alle violazioni gravi, ripetute o sistematiche degli obblighi di cui al Titolo II del decreto.

circostanze in larga parte fuori dal controllo dell'intermediario), ma per non aver adottato le cautele dovute nel processo di rilevazione e valutazione. Ne consegue che il giudizio di responsabilità è, in sostanza, un giudizio sull'adeguatezza del processo: se la funzione antiriciclaggio abbia impiegato strumenti idonei, se abbia valutato con diligenza, se abbia documentato le proprie determinazioni. Ed è precisamente per questo che la tracciabilità del percorso decisionale, di cui si è detto, assume rilievo non soltanto organizzativo ma anche difensivo.

La seconda è che l'imputazione presuppone un soggetto capace di rispondere. È un'osservazione che parrebbe ovvia, e che tuttavia diviene problematica quando una parte rilevante del processo è svolta da un sistema automatizzato. Un algoritmo, allo stato del diritto, non è un centro di imputazione: non è titolare di obblighi, non può essere sanzionato, non può essere destinatario di un rimprovero. Qualunque conseguenza giuridica derivi dal suo funzionamento deve dunque risalire a un soggetto umano o a un ente. Il problema, come si vedrà, è stabilire a quale.

3.2.2 Il delegato SOS tra automazione e responsabilità personale

La prima posizione da esaminare è quella del delegato alle segnalazioni, che dell'intero sistema è il punto di imputazione più immediato.

Le Istruzioni della UIF richiedono che tale soggetto sia dotato di idonei requisiti di indipendenza, autorevolezza e professionalità, e che svolga la propria attività con autonomia di giudizio, disponendo di tempo e di risorse adeguate[119]. La formula merita attenzione, perché l'autonomia di giudizio non è soltanto una garanzia contro le pressioni commerciali interne, ma diventa, nel contesto dell'automazione, qualcosa di più esigente: la capacità di discostarsi da ciò che il sistema suggerisce. Un delegato che non sia in condizione di contraddire l'algoritmo non esercita, in senso proprio, alcuna autonomia.

Qui si annida una difficoltà che l'ordinamento non ha ancora affrontato in modo esplicito. Come si è visto trattando dell'*automation bias*, la propensione a conformarsi all'output del sistema cresce con l'affidabilità percepita di quest'ultimo; e la possibilità di sindacarlo si riduce con la sua opacità. Il delegato si trova così stretto fra due esigenze in tensione: gli si chiede di validare criticamente ciò che, per costruzione, non è pienamente comprensibile. Se convalida senza comprendere, la sua validazione è formale e la responsabilità che continua a gravargli addosso è, in sostanza, una responsabilità per fatto altrui. Se rifiuta di convalidare ciò che non comprende, il sistema si blocca.

Vi è poi un profilo che merita di essere esplicitato, perché tocca la posizione personale di chi svolge questa funzione. L'adozione di un sistema di monitoraggio non è, di norma, una decisione del delegato: è una scelta compiuta dagli organi aziendali, spesso su impulso della funzione informatica e in dialogo con un fornitore esterno. Il delegato eredita quello strumento e ne subisce le caratteristiche, ma è lui a firmare la segnalazione. Si determina così una divaricazione fra chi sceglie il sistema e chi risponde delle decisioni

[119] UIF – Unità di Informazione Finanziaria per l'Italia, *Istruzioni per la rilevazione e la segnalazione delle operazioni sospette*, Provvedimento del 18 dicembre 2025, Parte Seconda, Sezione I, disponibile su: uif.bancaditalia.it.

che quel sistema orienta. È una divaricazione che la disciplina attuale non regola in modo specifico, e che rende tanto più importante il coinvolgimento della funzione antiriciclaggio nella fase di scelta della tecnologia: coinvolgimento che, non a caso, Banca d'Italia indica fra i tratti dei comportamenti più virtuosi, raccomandando che la funzione sia attivamente coinvolta nel processo di valutazione ex ante delle nuove soluzioni[120].

3.2.3 La responsabilità dell'intermediario e la scelta del sistema

Se dal singolo delegato si risale all'intermediario, il quadro si chiarisce sotto un profilo e si complica sotto un altro.

Si chiarisce perché, sul piano dei principi, non vi è dubbio che l'adozione di uno strumento tecnologico non trasferisca altrove la responsabilità. L'intermediario risponde dei propri presidi, quale che sia la tecnologia con cui li realizza; e la circostanza che l'inadempimento sia riconducibile a un malfunzionamento del sistema non vale, di per sé, a esonerarlo. Questo principio, per quanto intuitivo, ha una conseguenza operativa di rilievo: la scelta del sistema di monitoraggio diventa essa stessa oggetto di valutazione. Non è indifferente, sotto il profilo della diligenza, che l'intermediario abbia adottato uno strumento di cui è in grado di comprendere e governare il funzionamento oppure uno che, pur più performante, non gli consente di ricostruire le proprie determinazioni.

Torna qui, sotto una luce nuova, la critica dottrinale ricordata nel paragrafo precedente, secondo cui nei contesti a elevato impatto sarebbe preferibile impiegare modelli intrinsecamente interpretabili anziché affidarsi a spiegazioni costruite a posteriori. Ciò che in sede tecnica è una raccomandazione di buona pratica, in sede giuridica assume il valore di un criterio di diligenza: la preferenza per un modello opaco, quando fosse disponibile un'alternativa trasparente di prestazioni comparabili, è una scelta che l'intermediario deve essere in grado di giustificare.

Il quadro si complica, però, quando si considera che Banca d'Italia ha riscontrato, sul campo, esattamente le criticità che questa impostazione presuppone superate. L'indagine già richiamata segnala infatti fra i rischi quelli legati alla mancata comprensione dei meccanismi di funzionamento dei sistemi impiegati, e osserva che il ricorso a fornitori esterni specializzati può rendere difficoltosi i controlli su alcune fasi essenziali dei processi, fra cui il monitoraggio dell'operatività[121]. In altri termini: l'ordinamento pretende che l'intermediario governi il proprio sistema, ma la vigilanza constata che spesso non lo governa. Fra il dover essere e l'essere si apre uno spazio che, sul piano della responsabilità, è tutto a carico dell'intermediario.

3.2.4 Il fornitore fuori dal perimetro: il nodo dell'eterodirezione

Si giunge così al punto più problematico. Si è visto, trattando della relazione fra sistema e decisore umano, come i software di monitoraggio siano di regola progettati da soggetti estranei all'organizzazione dell'intermediario, e come ne discenda il rischio che la

[120] Banca d'Italia, *Indagine qualitativa sull'adozione di strumenti innovativi per l'adempimento degli obblighi AML/CFT*, 2 settembre 2025, par. 3, disponibile su: bancaditalia.it.
[121] Banca d'Italia, *Indagine qualitativa*, cit., par. 2, lett. b).

dottrina definisce di eterodirezione. Sul piano della responsabilità quel rilievo acquista un peso ulteriore. Nel disegno tradizionale dell'ordinamento bancario le regole si rivolgono a soggetti vigilati e presuppongono che le decisioni rilevanti siano assunte all'interno della loro organizzazione, da persone che ne condividono la soggezione ai controlli. Quando invece il criterio del giudizio è definito altrove, la classificazione delle *software-house* come meri fornitori di servizi ausiliari e strumentali risulta riduttiva, perché non coglie l'incidenza reale che la programmazione esercita sulle decisioni del soggetto vigilato[122]. La disciplina dell'esternalizzazione, concepita per rapporti nei quali il fornitore svolge un'attività che l'intermediario potrebbe svolgere da sé, mal si attaglia a una situazione nella quale il fornitore definisce i criteri stessi del giudizio.

Su questo fronte, peraltro, il quadro si è mosso dopo la riflessione appena richiamata. Il Regolamento (UE) 2022/2554 sulla resilienza operativa digitale del settore finanziario, noto come DORA e pienamente applicabile dal 17 gennaio 2025, interviene proprio sul rapporto fra entità finanziarie e fornitori terzi di servizi informatici: impone un quadro strutturato di gestione del rischio derivante da terzi, prescrive il contenuto minimo degli accordi contrattuali, richiede un registro delle forniture e piani di uscita, e obbliga a informare l'autorità competente prima di stipulare accordi per servizi a supporto di funzioni essenziali o importanti[123]. Soprattutto, per la prima volta il legislatore europeo spinge la vigilanza oltre il perimetro tradizionale: i fornitori designati come critici sono assoggettati a una sorveglianza diretta delle Autorità europee di vigilanza, dotate di poteri di ispezione, di richiesta di informazioni e di raccomandazione[124].

L'intervento attenua in parte la preoccupazione dottrinale, ma non la supera, e per ragioni che bisogna distinguere. DORA governa la resilienza operativa: si preoccupa che i sistemi funzionino, che gli incidenti siano gestiti, che la dipendenza da un fornitore non diventi un rischio sistemico. Non si occupa invece della qualità sostanziale della logica applicativa, sicché un modello di monitoraggio pienamente conforme può ciò nondimeno incorporare criteri di rilevazione opachi o inadeguati, senza che il regolamento offra strumenti per accorgersene. La sorveglianza diretta, inoltre, riguarda i soli fornitori designati come critici, categoria concepita per operatori di rilevanza sistemica quali i grandi fornitori di servizi cloud: la *software-house* che sviluppa l'applicativo antiriciclaggio di un intermediario di medie dimensioni ne resta con ogni probabilità fuori, vincolata soltanto per via contrattuale attraverso gli obblighi che l'entità finanziaria è tenuta a trasferire negli accordi. Resta infine fermo, e il regolamento lo conferma espressamente, che la responsabilità del rischio informatico grava sull'entità finanziaria e sui suoi organi di governo[125]: la disciplina rafforza gli obblighi di controllo dell'intermediario sul fornitore, ma non redistribuisce le conseguenze. Ciò che il

[122] LEMMA V., «Intelligenza Artificiale e sistemi di controllo», cit., in ordine al rischio di eterodirezione e ai limiti della qualificazione delle *software-house* quali fornitori di servizi ausiliari e strumentali.
[123] Regolamento (UE) 2022/2554 del Parlamento europeo e del Consiglio del 14 dicembre 2022, relativo alla resilienza operativa digitale per il settore finanziario, entrato in vigore il 16 gennaio 2023 e applicabile dal 17 gennaio 2025, Capo V e in particolare art. 28. Sul punto cfr. Banca d'Italia, *Comunicazione sull'applicazione del Regolamento DORA*, dicembre 2024, disponibile su: bancaditalia.it.
[124] Regolamento (UE) 2022/2554, artt. 31 ss., in ordine alla designazione dei fornitori terzi critici di servizi TIC e al relativo regime di sorveglianza.
[125] Regolamento (UE) 2022/2554, art. 5, che attribuisce all'organo di gestione dell'entità finanziaria la responsabilità ultima in materia di gestione dei rischi informatici.

legislatore ha disciplinato, in definitiva, è il rischio che il sistema si interrompa, non il rischio che il sistema sbagli.

Di qui la proposta, di ampio respiro, di estendere il perimetro del controllo pubblico anche alla produzione del software, così da evitare che lo sviluppo dell'intelligenza artificiale al di fuori di un'impresa vigilata si traduca in una fuga dalla supervisione[126]. È una prospettiva formulata prima che il legislatore europeo intervenisse in materia, e che proprio per questo merita di essere confrontata con quanto è accaduto dopo.

Occorre infatti registrare, sul punto, un dato di attualità che rende il problema più acuto anziché risolverlo. La Commissione europea aveva presentato nel 2022 una proposta di direttiva sulla responsabilità civile da intelligenza artificiale, destinata a integrare l'AI Act sul versante risarcitorio e ad affrontare le difficoltà probatorie che il danneggiato incontra di fronte a sistemi opachi. Tale proposta è stata ritirata: nel programma di lavoro per il 2025, adottato l'11 febbraio 2025, la Commissione ne ha disposto lo stralcio motivandolo con l'assenza di un accordo prevedibile, riservandosi di valutare se presentare un'altra proposta o scegliere un diverso approccio[127].

La conseguenza è che, allo stato, non esiste una disciplina europea armonizzata della responsabilità per i danni da sistemi di intelligenza artificiale. L'AI Act stabilisce obblighi di conformità, ma non disciplina in modo diretto la responsabilità civile; sicché l'attribuzione delle conseguenze resta affidata ai principi generali degli ordinamenti nazionali e alla disciplina sulla responsabilità per danno da prodotti difettosi, con le incertezze che ne derivano quando il prodotto è un modello statistico il cui malfunzionamento è difficile da provare e persino da definire[128]. La preoccupazione dottrinale, dunque, non è stata superata dall'intervento del legislatore: è rimasta, in larga misura, senza risposta.

3.2.5 Un problema di allocazione, non soltanto di imputazione

Le considerazioni sin qui svolte convergono verso una conclusione che va formulata con chiarezza.

Il problema che l'automazione pone all'antiriciclaggio non è propriamente quello di individuare un responsabile: sotto questo profilo l'ordinamento una risposta ce l'ha, ed è che risponde l'intermediario, e con esso chi vi svolge le funzioni rilevanti. Il problema è piuttosto che questa risposta, per quanto giuridicamente coerente, rischia di essere sostanzialmente iniqua e funzionalmente inefficace.

Iniqua, perché fa gravare le conseguenze su chi ha un controllo soltanto parziale sul processo. Il delegato risponde di una valutazione orientata da un sistema che non ha scelto e che spesso non è messo in condizione di comprendere. L'intermediario risponde del

[126] LEMMA V., «Intelligenza Artificiale e sistemi di controllo», cit.
[127] Commissione europea, *Programma di lavoro per il 2025*, adottato l'11 febbraio 2025, ove è disposto il ritiro della proposta di direttiva del Parlamento europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale, COM(2022) 496.
[128] Il riferimento è alla Direttiva (UE) 2024/2853 sulla responsabilità per danno da prodotti difettosi, che ha esteso la nozione di prodotto al software.

funzionamento di uno strumento la cui logica è stata definita da un fornitore che sfugge alla vigilanza e che, a sua volta, non risponde all'autorità.

Inefficace, perché il sistema sanzionatorio dovrebbe orientare i comportamenti verso una maggiore diligenza, ma per farlo occorre che gli incentivi raggiungano chi è nella posizione migliore per prevenire il danno. Se il produttore del software non risponde della qualità delle proprie scelte progettuali, l'incentivo a costruire sistemi comprensibili e verificabili resta debole; e il costo della loro opacità si scarica interamente a valle, su chi quei sistemi li utilizza.

Ne discende che il tema della responsabilità, nel campo che qui si esamina, è meno una questione di individuazione del soggetto e più una questione di allocazione del rischio lungo una catena che il diritto vigente considera ancora come una successione di rapporti bilaterali, mentre è divenuta una filiera. E ne discende, soprattutto, che la centralità del giudizio umano resta un principio fragile finché non è accompagnata dalle condizioni che lo rendono esercitabile: la comprensibilità dello strumento, la competenza di chi lo usa e la responsabilizzazione di chi lo costruisce. È su quest'ultimo terreno che si misura l'adeguatezza del quadro regolatorio emergente, oggetto del paragrafo che chiude il capitolo.

3.3 Riflessioni sull'AI Act

In questo paragrafo si procederà a verificare se il quadro regolatorio più recente — quello che il legislatore europeo ha costruito proprio per governare l'intelligenza artificiale — offra risposte adeguate sulle tematiche poste nei due paragrafi precedenti.

3.3.1 L'architettura dell'AI Act e il metodo del rischio

Il Regolamento (UE) 2024/1689, comunemente noto come AI Act, costituisce il primo tentativo al mondo di disciplinare in modo organico l'intelligenza artificiale. Entrato in vigore il 1° agosto 2024, esso non si limita a enunciare principi, ma costruisce un sistema di obblighi graduati secondo un criterio che merita attenzione perché condiziona tutto il ragionamento che segue: il regolamento non disciplina la tecnologia in quanto tale, bensì gli usi che se ne fanno, classificandoli in funzione del rischio che presentano per la sicurezza e per i diritti fondamentali[129].

Ne risulta una piramide a quattro livelli. Alla sommità si collocano le pratiche vietate, elencate all'articolo 5, che l'ordinamento considera incompatibili con i valori dell'Unione. Segue la categoria dei sistemi ad alto rischio, che non sono proibiti ma sono assoggettati a un regime di obblighi stringenti. Vi è poi una fascia di sistemi soggetti a soli obblighi di trasparenza, e infine tutto ciò che resta, sostanzialmente libero.

La distinzione interessante è quella che separa la seconda categoria dalle altre, perché è alla qualificazione come sistema ad alto rischio che l'AI Act ricollega gli obblighi più

[129] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale, in GUUE L 2024/1689 del 12 luglio 2024.

pregnanti. Fra questi figurano, in particolare, quelli che il capitolo ha individuato come cruciali: la trasparenza verso l'utilizzatore professionale, prevista dall'articolo 13, che impone che il sistema sia progettato in modo da garantire un funzionamento sufficientemente trasparente da consentire all'utilizzatore di comprenderne gli output e di impiegarli adeguatamente, corredandolo di istruzioni chiare e complete su caratteristiche, capacità e limiti delle prestazioni[130]; e la sorveglianza umana, disciplinata dall'articolo 14, che richiede che i sistemi siano progettati e sviluppati, anche mediante appropriati strumenti di interfaccia uomo-macchina, in modo da poter essere efficacemente supervisionati da persone fisiche durante il loro utilizzo[131].

Su quest'ultima previsione bisogna soffermarsi, perché coincide in modo notevole con quanto si è argomentato nel capitolo precedente. Le misure di sorveglianza devono essere proporzionate ai rischi, all'autonomia e al contesto d'uso del sistema. Le persone incaricate devono essere poste in condizione di monitorare, interpretare e disattendere l'output del sistema. Infine, la disciplina impone che si tenga conto della possibilità che si faccia eccessivo affidamento sugli output prodotti dall'intelligenza artificiale[132]. Il legislatore europeo ha recepito il problema *dell'automation bias* e ne ha fatto oggetto di una previsione normativa: non basta che un essere umano sia formalmente presente nel processo, occorre che sia messo nelle condizioni di esercitare un controllo reale e di discostarsi dal suggerimento della macchina.

Se applicati al monitoraggio antiriciclaggio, questi due articoli sembrerebbero fatti su misura per il problema che si è descritto. L'articolo 13 imporrebbe al fornitore del sistema di consegnare all'intermediario un prodotto comprensibile, corredato delle informazioni necessarie a interpretarne gli esiti; l'articolo 14 garantirebbe che il delegato conservi un potere effettivo di contraddire l'algoritmo. Il vuoto denunciato dalla dottrina parrebbe colmato. Occorre però verificare se quei due articoli si applichino davvero.

3.3.2 Il monitoraggio delle transazioni, in ambito AML, è un sistema ad alto rischio?

La risposta dipende dall'articolo 6 e dall'Allegato III, che individuano i sistemi rientranti nella categoria. L'Allegato elenca otto settori: biometria; infrastrutture critiche; istruzione e formazione professionale; occupazione e gestione dei lavoratori; accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali; attività di contrasto; migrazione, asilo e gestione delle frontiere; amministrazione della giustizia e processi democratici[133].

Scorrendo l'elenco, l'antiriciclaggio non compare. Nessuna delle otto voci menziona il monitoraggio delle transazioni finanziarie, la rilevazione delle anomalie operative o la segnalazione di operazioni sospette. Due categorie, tuttavia, meritano un esame più attento, perché a prima vista potrebbero sembrare comprensive del nostro caso.

[130] Regolamento (UE) 2024/1689, art. 13 (Trasparenza e fornitura di informazioni ai *deployer*).
[131] Regolamento (UE) 2024/1689, art. 14, par. 1 (Sorveglianza umana).
[132] Regolamento (UE) 2024/1689, art. 14.
[133] Regolamento (UE) 2024/1689, Allegato III, punti da 1 a 8.

La prima è quella dell'accesso ai servizi privati essenziali. Il punto 5, lettera b), qualifica come ad alto rischio i sistemi destinati a valutare l'affidabilità creditizia delle persone fisiche o a stabilirne il merito di credito. Si tratta della previsione più vicina al settore bancario, e potrebbe indurre a ritenere che ogni valutazione algoritmica della clientela vi rientri. La lettura del testo smentisce però questa impressione, e lo fa in modo esplicito: la disposizione fa salvi, escludendoli dalla categoria, i sistemi di intelligenza artificiale utilizzati allo scopo di individuare frodi finanziarie[134]. L'eccezione è significativa sotto due profili. Sul piano letterale, essa indica che il legislatore ha avuto presente l'ipotesi dei sistemi di rilevazione degli illeciti impiegati dagli intermediari e ha scelto, almeno in quella categoria, di non assoggettarli al regime rafforzato. Sul piano sistematico, essa colloca la finalità di individuazione degli illeciti finanziari fuori dal perimetro dell'alto rischio, in una logica che parrebbe voler evitare di gravare di oneri i presidi difensivi del sistema. Il monitoraggio antiriciclaggio, per finalità e struttura, è affine a quei sistemi antifrode: individua condotte anomale a tutela dell'integrità del sistema finanziario, non valuta l'accesso di un cliente a un servizio.

La seconda categoria potenzialmente rilevante è quella delle attività di contrasto, di cui al punto 6. Qui l'Allegato contempla, fra l'altro, i sistemi impiegati per valutare il rischio di commissione di un reato da parte di una persona fisica o per effettuarne la profilazione nel corso di indagini. La formulazione, però, circoscrive l'ambito in modo netto: le previsioni riguardano i sistemi destinati a essere utilizzati dalle autorità di contrasto o per loro conto, ovvero da istituzioni e organismi dell'Unione a loro sostegno[135]. Il riferimento è dunque all'apparato pubblico. Una simile qualificazione potrebbe astrattamente attagliarsi agli strumenti impiegati dalla UIF o dagli organi investigativi, ma non ai sistemi di un intermediario privato, che non è autorità di contrasto né agisce per suo conto, ma adempie a un obbligo di collaborazione.

Se ne deve concludere che i sistemi di *transaction monitoring* impiegati dai soggetti obbligati non sono, allo stato, qualificati come ad alto rischio dall'AI Act. Ne consegue che gli obblighi degli articoli 13 e 14 (trasparenza verso l'utilizzatore e sorveglianza umana) non si applicano loro direttamente. La conclusione va precisata con due avvertenze. La prima è che la qualificazione dipende dalla finalità concreta del sistema: uno strumento che, oltre a rilevare anomalie, fosse impiegato per decidere dell'accesso del cliente a servizi essenziali potrebbe ricadere altrove nell'elenco. La seconda è che l'Allegato III è modificabile dalla Commissione ai sensi dell'articolo 7, sicché la situazione non è immutabile.

3.3.3 Il diritto alla spiegazione e la sua inapplicabilità al segnalato

L'esclusione dal perimetro dell'alto rischio produce una conseguenza ulteriore, che merita un esame autonomo perché tocca il tema con cui questo capitolo si è aperto: la spiegabilità.

---

[134] Regolamento (UE) 2024/1689, Allegato III, punto 5, lett. b), dove sono qualificati ad alto rischio i sistemi destinati a valutare l'affidabilità creditizia delle persone fisiche o a stabilirne il merito di credito, «a eccezione dei sistemi di IA utilizzati allo scopo di individuare frodi finanziarie».

[135] Regolamento (UE) 2024/1689, Allegato III, punto 6, lett. d) ed e).

Fra le disposizioni dell'AI Act ve n'è una che riguarda specificamente la posizione di chi subisce una decisione assistita da un algoritmo. L'articolo 86 riconosce un diritto alla spiegazione dei singoli processi decisionali. La persona interessata può chiedere all'utilizzatore professionale spiegazioni chiare e significative su due punti: quale ruolo abbia avuto il sistema nella procedura decisionale e quali siano stati i principali elementi della decisione adottata. Il diritto opera però entro confini precisi. Occorre anzitutto che la decisione si fondi sull'output di un sistema ad alto rischio fra quelli elencati nell'Allegato III, con esclusione di quelli relativi alle infrastrutture critiche di cui al punto 2. Occorre inoltre che la decisione produca effetti giuridici, o comunque incida in modo significativo sull'interessato, il quale deve ritenerne pregiudicati la propria salute, sicurezza o i propri diritti fondamentali[136].

La disposizione ha una fisionomia precisa. Non attribuisce un diritto di accesso al funzionamento tecnico del sistema, ma alla comprensibilità della decisione: si tratta di sapere come il sistema sia entrato nel processo e quali fattori principali abbiano condotto all'esito, non di ottenere la divulgazione dell'algoritmo. Ed è un diritto che l'interessato esercita su richiesta, non un obbligo che l'utilizzatore debba adempiere d'ufficio.

Trasposta al nostro campo, la previsione sarebbe di grande interesse. La segnalazione di operazione sospetta è precisamente una decisione che incide significativamente sulla posizione di una persona: attiva l'analisi finanziaria dell'Unità di Informazione Finanziaria, può condurre alla trasmissione agli organi investigativi e, di lì, all'apertura di accertamenti. Un diritto alla spiegazione, in tale contesto, colmerebbe un'asimmetria evidente.

Senonché l'articolo 86 condivide il presupposto applicativo degli articoli 13 e 14: si riferisce alle decisioni fondate sull'output di un sistema ad alto rischio elencato nell'Allegato III. Escluso il monitoraggio antiriciclaggio da quella categoria, il diritto alla spiegazione non trova applicazione. Il cliente segnalato non può invocarlo.

L'analisi non può però fermarsi all'AI Act, perché il diritto alla spiegazione non nasce con esso. Il regolamento generale sulla protezione dei dati contiene, all'articolo 22, una disciplina delle decisioni basate unicamente su trattamenti automatizzati che producano effetti giuridici o incidano in modo analogamente significativo sull'interessato, riconoscendo a quest'ultimo il diritto di ottenere l'intervento umano, di esprimere la propria opinione e di contestare la decisione[137]. Si tratta di una tutela autonoma, applicabile indipendentemente dalla qualificazione del sistema come ad alto rischio, e che lo stesso articolo 86 dell'AI Act fa salva, precisando di applicarsi solo nella misura in cui il diritto non sia altrimenti previsto dal diritto dell'Unione[138].

Di particolare interesse è l'interpretazione che di tale disciplina ha dato la Corte di giustizia nella sentenza SCHUFA. Chiamata a pronunciarsi sul punteggio di affidabilità creditizia elaborato da una società di informazioni commerciali, la Corte ha stabilito che il calcolo automatizzato di un tasso di probabilità relativo alla capacità di una persona di

[136] Regolamento (UE) 2024/1689, art. 86, par. 1 (Diritto alla spiegazione dei singoli processi decisionali).
[137] Regolamento (UE) 2016/679 (GDPR), art. 22, in materia di processo decisionale automatizzato relativo alle persone fisiche.
[138] Regolamento (UE) 2024/1689, art. 86, par. 3.

onorare gli impegni di pagamento costituisce un processo decisionale automatizzato ai sensi dell'articolo 22, quando il destinatario del punteggio vi si conformi in modo determinante nell'assumere la propria decisione[139]. Il criterio ha rilievo generale e tocca da vicino il problema esaminato in questo capitolo: ciò che qualifica una decisione come automatizzata non è la presenza formale di un operatore umano nel processo, bensì il peso effettivo che l'output del sistema esercita sull'esito. Una validazione che si risolva nella ratifica di un punteggio già calcolato non sottrae la decisione alla disciplina dell'articolo 22; semmai conferma che essa è automatizzata nella sostanza.

L'applicazione di questo apparato al settore antiriciclaggio incontra però il medesimo ostacolo già visto a proposito dell'AI Act, e per la stessa ragione. Il regolamento sulla protezione dei dati ammette che il diritto dell'Unione e degli Stati membri limiti la portata dei diritti dell'interessato per finalità di interesse pubblico rilevante, fra le quali rientra la prevenzione del riciclaggio[140].

A ben vedere, tuttavia, il punto merita una precisazione che ne muta il senso, e che conviene esplicitare per non incorrere in un fraintendimento. Anche a voler ipotizzare la qualificazione del sistema come ad alto rischio, il diritto alla spiegazione si scontrerebbe nel nostro settore con un ostacolo di ordine diverso e assai più radicale: il divieto di comunicazione previsto dall'articolo 39 del decreto antiriciclaggio, che vieta ai soggetti obbligati di informare il cliente dell'avvenuta segnalazione[141]. Il segnalato non soltanto non ha diritto a una spiegazione: non deve nemmeno sapere che una decisione è stata adottata nei suoi confronti. La stessa disposizione europea, del resto, prevede che il diritto non si applichi dove sussistano eccezioni o limitazioni previste dal diritto dell'Unione o nazionale[142], e il divieto di *tipping off* è precisamente una di tali limitazioni, funzionale a preservare l'efficacia delle indagini.

Ne discende una considerazione che riguarda l'intera architettura del sistema. Nell'antiriciclaggio la trasparenza verso l'interessato non è soltanto assente: è vietata, e per ragioni che l'ordinamento considera prevalenti. Questa scelta, del tutto comprensibile sul piano della politica criminale, ha però un corollario che è bene rendere esplicito: se il destinatario della decisione non può essere informato né contestarla, l'intero peso della garanzia si sposta a monte, sui controlli interni all'intermediario e sulla vigilanza pubblica. La qualità della motivazione, di cui si è discusso all'inizio del capitolo, non serve dunque a permettere al segnalato di difendersi (cosa che gli è preclusa), ma a consentire all'autorità di verificare che la segnalazione sia fondata. È una garanzia che opera per interposta persona, il che la rende tanto più fragile quanto meno l'autorità è in condizione di ricostruire il percorso che ha portato alla segnalazione. E si torna, per questa

[139] Corte di giustizia dell'Unione europea, 7 dicembre 2023, causa C-634/21, *SCHUFA Holding*, ECLI:EU:C:2023:957.
[140] Regolamento (UE) 2016/679, art. 23, in ordine alle limitazioni ai diritti dell'interessato.
[141] D.Lgs. 21 novembre 2007, n. 231, art. 39, in materia di divieto di comunicazione dell'avvenuta segnalazione.
[142] Regolamento (UE) 2024/1689, art. 86, par. 2, dove si prevede che il diritto alla spiegazione non si applichi all'uso di sistemi di intelligenza artificiale per i quali sono previste eccezioni o limitazioni in virtù del diritto dell'Unione o del diritto nazionale.

via, al problema della spiegabilità: in un sistema nel quale il controllato non può contestare, la comprensibilità del processo è l'unico presidio residuo.

3.3.4 Il rinvio degli obblighi e lo stato attuale

Alla constatazione appena svolta si aggiunge un dato di cronaca normativa che accentua l'incertezza del quadro. Il calendario applicativo dell'AI Act, che prevedeva l'applicazione degli obblighi sui sistemi ad alto rischio a partire dal 2 agosto 2026, è stato modificato. Con il Regolamento (UE) 2026/1744, noto come Omnibus digitale sull'intelligenza artificiale approvato dal Parlamento europeo il 16 giugno 2026, adottato dal Consiglio il 29 giugno ed entrato in vigore il 27 luglio dello stesso anno; gli obblighi relativi ai sistemi ad alto rischio elencati nell'Allegato III sono stati posticipati al 2 dicembre 2027, mentre quelli concernenti i sistemi integrati come componenti di sicurezza in prodotti già soggetti alla normativa di armonizzazione slittano al 2 agosto 2028[143].

Il rinvio è stato motivato con la necessità di completare gli strumenti tecnici e di governance indispensabili all'applicazione, e non incide sull'impianto del regolamento, che resta fondato sull'approccio basato sul rischio; restano peraltro applicabili le disposizioni già in vigore, fra cui i divieti e gli obblighi di trasparenza[144]. Ai fini di questa trattazione, tuttavia, il dato rileva soprattutto perché conferma che il quadro europeo sull'intelligenza artificiale è ancora in assestamento, e che i presidi più incisivi — quelli che avrebbero potuto offrire un modello anche per il settore che qui si esamina — non hanno ancora dispiegato i propri effetti.

3.3.5 Un paradosso e le sue implicazioni

Le conclusioni raggiunte compongono un quadro che merita di essere guardato nel suo insieme, perché ne emerge un paradosso.

L'AI Act sottopone a obblighi stringenti di trasparenza e sorveglianza umana i sistemi che valutano il merito creditizio di una persona, selezionano un candidato a un posto di lavoro o determinano l'ammissione a un corso di studi, e riconosce a chi ne subisce gli effetti un diritto alla spiegazione. Non applica invece nulla di tutto ciò ai sistemi che, individuando un'anomalia nell'operatività di un cliente, possono innescare una segnalazione destinata a confluire nell'analisi dell'Unità di Informazione Finanziaria e, di lì, in un'indagine della Guardia di Finanza. È difficile sostenere che la seconda ipotesi incida sulla posizione della persona meno della prima: chi si vede negare un finanziamento lo sa, e può reagire; il segnalato ignora tutto.

Il paradosso si spiega, probabilmente, con la ratio dell'esclusione: gravare di oneri i sistemi antifrode e antiriciclaggio avrebbe potuto disincentivarne l'adozione, indebolendo presidi che l'ordinamento stesso impone; e, come si è visto, la riservatezza è connaturata al meccanismo delle segnalazioni. È una scelta di politica del diritto comprensibile, che

[143] Regolamento (UE) 2026/1744 del Parlamento europeo e del Consiglio dell'8 luglio 2026, che modifica i regolamenti (UE) 2024/1689, (UE) 2018/1139 e (UE) 2023/1230, in GUUE, serie L, del 24 luglio 2026.
[144] *Ibidem*.

sacrifica le garanzie individuali all'efficacia del controllo. Resta però il fatto che la scelta produce, come effetto, l'esclusione dal regime di garanzia proprio di quei sistemi che operano nel silenzio e senza contraddittorio.

Ne discende una conseguenza pratica di rilievo. Il presidio dell'intervento umano, nel monitoraggio antiriciclaggio, non poggia dunque sull'articolo 14 del regolamento europeo, ma sulla disposizione delle Istruzioni della UIF già esaminata, che subordina l'impiego di sistemi di intelligenza artificiale alla validazione umana delle anomalie da essi evidenziate[145]. È significativo che la stessa disposizione richiami, in nota, il Regolamento (UE) 2024/1689: il rinvio suona come un'apertura prudente, che non presuppone l'applicabilità automatica del regolamento ma la rimette alla verifica caso per caso.

Occorre però prendere atto della differenza di rango e di ampiezza fra le due fonti. L'articolo 14 dell'AI Act si rivolge al fornitore e gli impone di progettare il sistema in modo che sia sorvegliabile, con obblighi presidiati da un apparato sanzionatorio europeo. La disposizione della UIF si rivolge invece al destinatario degli obblighi antiriciclaggio e gli impone di accompagnare l'uso del sistema con valutazioni umane: prescrive cioè un comportamento a chi il sistema lo utilizza, non a chi lo costruisce. Il che riporta esattamente al nodo esaminato nel paragrafo precedente: l'obbligo grava su chi non ha progettato lo strumento, mentre chi lo ha progettato resta fuori dal perimetro.

Si conferma così, e da ultimo, la diagnosi formulata in dottrina prima che il legislatore europeo intervenisse. L'auspicio era che il controllo pubblico si estendesse alla produzione del software, per evitare che lo sviluppo dell'intelligenza artificiale al di fuori di un'impresa vigilata si traducesse in una fuga dalla supervisione[146]. L'AI Act ha effettivamente compiuto quel passo, imponendo obblighi ai fornitori e non soltanto agli utilizzatori: ma lo ha fatto per i sistemi che ha qualificato ad alto rischio, categoria dalla quale il monitoraggio antiriciclaggio resta fuori. Sommato al ritiro della proposta di direttiva sulla responsabilità civile, il risultato è che, proprio nel settore in cui l'ordinamento pretende con maggior rigore che la decisione resti umana e motivata, la disciplina degli strumenti che quella decisione orientano è la più esile.

[145] UIF – Unità di Informazione Finanziaria per l'Italia, *Istruzioni per la rilevazione e la segnalazione delle operazioni sospette*, Provvedimento del 18 dicembre 2025, Parte Prima, Sezione II, disponibile su: uif.bancaditalia.it.

[146] LEMMA V., «Intelligenza Artificiale e sistemi di controllo: quali prospettive regolamentari?», in *Rivista Trimestrale di Diritto dell'Economia*, Supplemento n. 2 al n. 3/2021.

Conclusioni

Il percorso compiuto in questa tesi ha preso le mosse da una domanda che, a distanza di tre capitoli, può essere ripresa nella sua formulazione iniziale: nell'ambito del monitoraggio delle transazioni e delle segnalazioni di operazioni sospette, qual è e quale deve essere il ruolo del fattore umano, ora che l'intelligenza artificiale promette di trasformare i processi di individuazione dell'anomalia?

La risposta che l'indagine consente di formulare è netta nella sostanza e articolata nelle implicazioni. L'intelligenza artificiale non può sostituire il giudizio umano nel processo di segnalazione, e non lo può per ragioni che non sono soltanto tecniche ma strutturalmente giuridiche. Ciò non significa, però, che la sua adozione sia neutrale rispetto al ruolo dell'uomo: al contrario, essa lo ridefinisce profondamente, e nel farlo mette alla prova la tenuta di alcuni presupposti su cui l'intero sistema di prevenzione è costruito.

Il primo capitolo ha ricostruito l'architettura del sistema antiriciclaggio italiano, muovendo dal decreto antiriciclaggio e giungendo alle Istruzioni della UIF del dicembre 2025, entrate in vigore proprio mentre questo lavoro veniva redatto; soffermandosi inoltre sui regolamenti e direttive europei in tema (pacchetto AML). Ne è emerso un impianto normativo ben strutturato, che assegna agli intermediari un ruolo di presidio attivo e che affida alla soglia del sospetto ragionevole il compito di selezionare ciò che merita di essere portato all'attenzione dell'autorità. Ma ne sono emerse anche disfunzioni che il tempo ha reso strutturali: un volume di segnalazioni che nel 2025 ha superato le centoquarantacinquemila unità[147], una quota di *alert* infondati che nei sistemi tradizionali supera il novanta per cento e una tendenza alla segnalazione difensiva.

Il secondo capitolo ha esaminato la risposta tecnologica a queste inefficienze. L'analisi dei modelli di apprendimento automatico ha mostrato potenzialità concrete: la capacità di apprendere dagli esiti passati, quella di riconoscere schemi relazionali che sfuggono a qualunque regola statica, quella di adattare la nozione di normalità al profilo del singolo cliente. Il caso del modello sperimentato dalla stessa Unità di Informazione Finanziaria per l'identificazione delle imprese collegate alla criminalità organizzata ne offre una dimostrazione istituzionale, con risultati apprezzabili anche se lontani dalla perfezione. Ma il medesimo capitolo ha imposto due correttivi all'entusiasmo. Il primo viene dai dati: l'indagine condotta dalla Banca d'Italia nel 2025 rivela che, fra le banche già orientate all'innovazione, soltanto un quinto ha adottato tecnologie innovative nel monitoraggio transazionale, e che il grado di diffusione del *machine learning* in questo ambito resta basso[148]; sicché la trasformazione di cui si discute è, in larga misura, ancora attesa piuttosto che compiuta. Il secondo viene dalla natura stessa dei modelli: quelli più efficaci tendono a essere anche i meno intelligibili, con un divario che nei sistemi più complessi non è colmabile a posteriori.

[147] UIF – Unità di Informazione Finanziaria per l'Italia, *Rapporto annuale per il 2025*, n. 18, 2026, disponibile su: uif.bancaditalia.it, per il dato relativo al volume delle segnalazioni ricevute.
[148] Banca d'Italia, *Indagine qualitativa sull'adozione di strumenti innovativi per l'adempimento degli obblighi AML/CFT*, 2 settembre 2025, disponibile su: bancaditalia.it.

Il terzo capitolo ha portato questa tensione sul terreno giuridico, ed è qui che il lavoro ha raggiunto i suoi esiti meno scontati. Si è visto che l'ordinamento antiriciclaggio pretende una motivazione: le Istruzioni della UIF richiedono che il segnalante esponga il processo logico-deduttivo seguito, e vietano espressamente la motivazione per mero rinvio agli indicatori di anomalia. Si è visto poi che le tecniche di intelligenza artificiale spiegabile, per quanto utili, non restituiscono il ragionamento del sistema (che in senso proprio non esiste) ma un'approssimazione costruita dall'esterno, instabile, talora infedele e persino manipolabile. Si è visto infine che la responsabilità continua a gravare su chi utilizza lo strumento, mentre chi lo progetta resta fuori dal perimetro della vigilanza. L'esito forse più significativo dell'indagine riguarda però l'AI Act. La lettura dell'Allegato III ha mostrato che i sistemi di monitoraggio antiriciclaggio impiegati dagli intermediari non rientrano fra quelli qualificati ad alto rischio: la disposizione più prossima, relativa alla valutazione del merito creditizio, esclude testualmente i sistemi destinati a individuare frodi finanziarie, e le previsioni sulle attività di contrasto riguardano gli strumenti delle autorità pubbliche, non quelli dei privati che con esse collaborano[149]. Ne consegue che gli obblighi di trasparenza e di sorveglianza umana previsti dal regolamento non si applicano direttamente a questi sistemi, e che nemmeno il diritto alla spiegazione riconosciuto dall'articolo 86 può essere invocato da chi ne subisce gli effetti, il quale in forza del divieto di comunicazione non può nemmeno sapere di essere stato segnalato.

Da questo insieme di rilievi si può ora rispondere alla domanda iniziale, e la risposta si articola su tre livelli.

Sul piano giuridico, il giudizio umano è insostituibile perché l'ordinamento non conosce, allo stato, alcun centro di imputazione diverso dalla persona. La segnalazione è un atto che impegna responsabilità amministrative ed eventualmente penali, e tali responsabilità presuppongono un soggetto capace di rispondere. Un algoritmo non è titolare di obblighi, non può essere sanzionato, non può essere destinatario di un rimprovero. Delegargli la decisione significherebbe creare un vuoto che l'ordinamento non tollera. È una constatazione che vale indipendentemente dal grado di accuratezza raggiunto dai modelli: nessun progresso tecnico può, di per sé, produrre un soggetto giuridico.

Sul piano funzionale, il giudizio umano è insostituibile perché ciò che il sistema produce e ciò che l'ordinamento richiede appartengono a due ordini distinti. L'algoritmo restituisce un'evidenza statistica: questa operatività si discosta dal modello atteso. La norma pretende una valutazione: questa operatività non trova giustificazione lecita alla luce di quanto si conosce del cliente. Il passaggio dall'una all'altra non è automatico né automatizzabile, perché richiede di contestualizzare, di interpretare le giustificazioni addotte, di valutare la plausibilità economica di una condotta. È esattamente ciò che le Istruzioni della UIF domandano quando esigono che la motivazione circoscriva l'anomalia rispetto al profilo concreto del soggetto.

---

[149] Regolamento (UE) 2024/1689, Allegato III, punto 5, lett. b), e punto 6. L'Allegato è peraltro modificabile dalla Commissione mediante atti delegati, ai sensi dell'art. 7.

Sul piano organizzativo, infine, l'automazione non riduce il peso del fattore umano ma lo accresce, spostandolo. L'analista sollevato dalla verifica massiva di allarmi infondati è chiamato a esercitare competenze di grado superiore: interpretare, contestualizzare, sindacare l'output del sistema. Il rischio, in questa configurazione, non è che la macchina renda superfluo l'uomo, ma che l'uomo abdichi al proprio ruolo per eccesso di fiducia. È il fenomeno *dell'automation bias*, che si presenta con una struttura paradossale: quanto più il sistema è affidabile, tanto più forte è la propensione a fidarsene, e tanto meno probabile è che l'operatore colga l'errore proprio quando esso si verifica.

Sarebbe però riduttivo concludere che il principio della centralità umana renda il quadro soddisfacente. L'indagine ha mostrato l'opposto: quel principio è affermato con chiarezza, ma le condizioni che ne consentono l'esercizio effettivo sono in larga parte assenti.

La prima condizione è la comprensibilità. Un delegato può validare soltanto ciò che è in grado di comprendere; se il sistema non gli offre alcun appiglio, la sua validazione diventa ratifica e la motivazione che ne segue una razionalizzazione a posteriori, costruita per giustificare una conclusione raggiunta altrove. Non è un rischio astratto: la stessa Banca d'Italia annovera fra le criticità riscontrate quelle legate alla mancata comprensione dei meccanismi di funzionamento dei sistemi impiegati.

La seconda condizione è la responsabilizzazione di chi il sistema lo costruisce. Finché il produttore del software non risponde delle proprie scelte progettuali, l'incentivo a realizzare strumenti comprensibili resta debole, e il costo dell'opacità si scarica interamente su chi li utilizza. Il risultato è un'allocazione del rischio che appare al tempo stesso iniqua e inefficiente, perché non raggiunge chi sarebbe nella posizione migliore per prevenire il danno.

La terza condizione è la coerenza del quadro regolatorio. E qui si registra il paradosso più netto emerso da questa ricerca: l'ordinamento europeo protegge con maggiore intensità chi si vede negare un prestito rispetto a chi viene segnalato all'autorità antiriciclaggio. Il primo sa, può chiedere spiegazioni, può contestare; il secondo ignora tutto, e il sistema che lo ha individuato non è soggetto ad alcuno degli obblighi rafforzati previsti per l'alto rischio. La scelta ha una sua logica: non gravare di oneri i presidi difensivi e preservare la riservatezza delle indagini.

Da queste considerazioni discendono alcune prospettive.

La prima riguarda la scelta dei modelli. Se l'opacità non è una fatalità ma la conseguenza di una scelta progettuale, allora la preferenza per un sistema interpretabile cessa di essere una raccomandazione tecnica e diventa un criterio di diligenza. L'intermediario che adotti uno strumento di cui non è in grado di ricostruire le determinazioni compie una scelta che deve poter giustificare.

La seconda riguarda il perimetro della vigilanza. La proposta, formulata in dottrina, di estendere il controllo pubblico alla produzione del software conserva intatta la propria attualità. L'AI Act ha mostrato che imporre obblighi ai fornitori è possibile: resta da

decidere se il monitoraggio antiriciclaggio meriti di rientrare fra i sistemi per i quali tali obblighi valgono.

La terza riguarda la competenza. Nessuna previsione normativa sulla sorveglianza umana produce effetti se chi è chiamato a esercitarla non dispone degli strumenti per farlo. La formazione delle funzioni antiriciclaggio (non soltanto giuridica, ma anche relativa al funzionamento e ai limiti dei modelli statistici) diventa in questa prospettiva un presidio sostanziale, e non un adempimento accessorio.

Vi è, infine, una considerazione di ordine più generale. Questa ricerca ha mostrato che l'intelligenza artificiale, nel campo esaminato, non risolve i problemi: li sposta. Dove prima l'intermediario doveva gestire l'inefficienza di un sistema che produceva troppi allarmi, ora deve gestire la complessità e l'opacità di un sistema che ne produce meno ma di cui comprende meno il funzionamento. È uno spostamento che vale la pena compiere, perché il guadagno in efficacia è reale; ma è uno spostamento, non una soluzione, e trattarlo come tale sarebbe l'errore più grave.

La segnalazione di operazione sospetta resta, in definitiva, ciò che il legislatore ha voluto che fosse: un atto di valutazione, fondato su ragioni esprimibili e imputabile a chi lo compie. La tecnologia può renderlo più informato, più tempestivo, meno dispersivo. Non può renderlo automatico, perché ciò che vi si esercita non è un calcolo di probabilità ma un giudizio — e il giudizio, per definizione, appartiene a chi ne risponde.

Bibliografia

Monografie e contributi dottrinali

LANNOO K. – PARLOUR R., *Anti-Money Laundering in the EU: Time to get serious*, CEPS-ECRI Task Force Report, Centre for European Policy Studies, Bruxelles, gennaio 2021.

RAZZANTE R., *Manuale di legislazione e prassi dell'antiriciclaggio*, II ed., Giappichelli, Torino, 2023.

Articoli e contributi in riviste e atti di convegno

AJAGBE S.A. – MAJOLA S. – MUDALI P., «Comparative analysis of machine learning algorithms for money laundering detection», in *Discover Artificial Intelligence*, vol. 5, 2025, art. 144.

ALVAREZ-MELIS D. – JAAKKOLA T.S., «On the Robustness of Interpretability Methods», presentato al 2018 ICML Workshop on Human Interpretability in Machine Learning (WHI 2018), Stoccolma, 2018, arXiv:1806.08049.

CARIELLO P. – DE SIMONI M. – IEZZI S., «Un modello di *machine learning* per l'identificazione di aziende collegate alla criminalità organizzata in Italia», in UIF, *Quaderni dell'antiriciclaggio – Analisi e studi*, n. 22, giugno 2024.

DALLA PELLEGRINA L. – DI MAIO G. – MASCIANDARO D. – SARACENO M., «Are Bankers "Crying Wolf"? Type I, Type II Errors and Deterrence in Anti-Money Laundering: The Italian Case», in *Italian Economic Journal*, vol. 9, n. 2, 2023, pp. 587-615.

GARA M. – PAUSELLI C., «Looking at 'Crying Wolf' from a Different Perspective: An Attempt at Detecting Banks Under- and Over-Reporting of Suspicious Transactions», in *Italian Economic Journal*, vol. 6, n. 2, 2020, pp. 299-324 (già in UIF, *Quaderni dell'antiriciclaggio – Analisi e studi*, n. 4, novembre 2015).

KUTE D.V. – PRADHAN B. – SHUKLA N. – ALAMRI A., «Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering – A Critical Review», in *IEEE Access*, vol. 9, 2021, pp. 82300-82317.

LEMMA V., «Intelligenza Artificiale e sistemi di controllo: quali prospettive regolamentari?», in *Rivista Trimestrale di Diritto dell'Economia*, Supplemento n. 2 al n. 3/2021, pp. 319-342.

LIBERTI J.M. – PETERSEN M.A., «Information: Hard and Soft», in *The Review of Corporate Finance Studies*, vol. 8, n. 1, 2019, pp. 1-41.

LUNDBERG S.M. – LEE S.I., «A Unified Approach to Interpreting Model Predictions», in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4765-4774.

NGUYEN V.D. – NGUYEN K.H., «Enhancing Money Laundering Detection: A Comparative Study of Machine Learning Techniques and Sampling Methods», in *Computational Science and Its Applications – ICCSA 2025*, Lecture Notes in Computer Science, vol. 15650, Springer, Cham, 2025, pp. 150-161.

OZTAS B. – CETINKAYA D. – ADEDOYIN F. – BUDKA M. – AKSU G. – DOGAN H., «Transaction Monitoring in Anti-Money Laundering: A Qualitative Analysis and Points of View from Industry», in *Future Generation Computer Systems*, vol. 159, 2024, pp. 161-171.

RIBEIRO M.T. – SINGH S. – GUESTRIN C., «"Why Should I Trust You?": Explaining the Predictions of Any Classifier», in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, 2016, pp. 1135-1144.

RUDIN C., «Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead», in *Nature Machine Intelligence*, vol. 1, n. 5, 2019, pp. 206-215.

SLACK D. – HILGARD S. – JIA E. – SINGH S. – LAKKARAJU H., «Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods», in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, ACM, New York, 2020, pp. 180-186.

TAKÁTS E., «A Theory of "Crying Wolf": The Economics of Money Laundering Enforcement», in *The Journal of Law, Economics, and Organization*, vol. 27, n. 1, 2011, pp. 32-78.

UNGER B. – VAN WAARDEN F., «How to Dodge Drowning in Data? Rule- and Risk-Based Anti Money Laundering Policies Compared», in *Review of Law & Economics*, vol. 5, n. 2, 2009, pp. 953-985.

VISANI G. – BAGLI E. – CHESANI F. – POLUZZI A. – CAPUZZO D., «Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models», in *Journal of the Operational Research Society*, vol. 73, n. 1, 2022, pp. 91-101.

WEBER M. – DOMENICONI G. – CHEN J. – WEIDELE D.K.I. – BELLEI C. – ROBINSON T. – LEISERSON C.E., «Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics», presentato all'Anomaly Detection in Finance Workshop, 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Anchorage, 2019, arXiv:1908.02591.

ZETZSCHE D.A. – ARNER D.W. – BUCKLEY R.P. – TANG B.W., «Artificial Intelligence in Finance: Putting the Human in the Loop», *University of Hong Kong Faculty of Law Research Paper* n. 2020/006, febbraio 2020.

Documenti istituzionali

BANCA D'ITALIA, *Regolamento (UE) 2022/2554 del Parlamento europeo e del Consiglio del 14 dicembre 2022 relativo alla resilienza operativa digitale per il settore finanziario (Regolamento DORA). Comunicazione della Banca d'Italia*, dicembre 2024, disponibile su: bancaditalia.it.

BANCA D'ITALIA – UNITÀ DI SUPERVISIONE E NORMATIVA ANTIRICICLAGGIO, *Indagine qualitativa sull'adozione di strumenti innovativi per l'adempimento degli obblighi AML/CFT*, 2 settembre 2025, disponibile su: bancaditalia.it.

COMMISSIONE EUROPEA, *Programma di lavoro della Commissione per il 2025*, COM(2025) 45 final, 11 febbraio 2025, Allegato IV.

GAFI – GRUPPO D'AZIONE FINANZIARIA INTERNAZIONALE, *Guidance for a Risk-Based Approach – The Banking Sector*, FATF/OECD, Parigi, ottobre 2014, disponibile su: fatf-gafi.org.

PRINCIPIO DI REVISIONE INTERNAZIONALE (ISA Italia) 200, *Obiettivi generali del revisore indipendente e svolgimento della revisione contabile in conformità ai principi di revisione internazionali (ISA Italia)*.

UIF – UNITÀ DI INFORMAZIONE FINANZIARIA PER L'ITALIA, *Quaderni dell'antiriciclaggio – Dati statistici*, II semestre 2016, disponibile su: uif.bancaditalia.it.

UIF, *Newsletter*, n. 1, 2019, allegato statistico, disponibile su: uif.bancaditalia.it.

UIF, *Schemi rappresentativi di comportamenti anomali – Operatività connessa con illeciti fiscali*, 10 novembre 2020, disponibile su: uif.bancaditalia.it.

UIF, *Newsletter*, n. 1, 2024, allegato statistico, disponibile su: uif.bancaditalia.it.

UIF, *Rapporto annuale per il 2023*, n. 16, 2024, disponibile su: uif.bancaditalia.it.

UIF, *Rapporto annuale per il 2024*, n. 17, 2025, disponibile su: uif.bancaditalia.it.

UIF, *Rapporto annuale per il 2025*, n. 18, 2026, disponibile su: uif.bancaditalia.it.

UIF, *Operatività connessa con truffe, frodi agevolate dalla tecnologia, money muling e altri reati informatici*, Comunicazione dell'8 giugno 2026, disponibile su: uif.bancaditalia.it.

Normativa di riferimento

Normativa dell'Unione europea

Direttiva 2001/97/CE del Parlamento europeo e del Consiglio, del 4 dicembre 2001, recante modifica della direttiva 91/308/CEE del Consiglio relativa alla prevenzione dell'uso del sistema finanziario a scopo di riciclaggio dei proventi di attività illecite (II Direttiva antiriciclaggio).

Direttiva 2005/60/CE del Parlamento europeo e del Consiglio, del 26 ottobre 2005, relativa alla prevenzione dell'uso del sistema finanziario a scopo di riciclaggio dei proventi di attività criminose e di finanziamento del terrorismo (III Direttiva antiriciclaggio).

Direttiva (UE) 2015/849 del Parlamento europeo e del Consiglio, del 20 maggio 2015, relativa alla prevenzione dell'uso del sistema finanziario a fini di riciclaggio o finanziamento del terrorismo (IV Direttiva antiriciclaggio).

Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati (regolamento generale sulla protezione dei dati).

Direttiva (UE) 2018/843 del Parlamento europeo e del Consiglio, del 30 maggio 2018, che modifica la direttiva (UE) 2015/849 (V Direttiva antiriciclaggio).

Regolamento (UE) 2022/2554 del Parlamento europeo e del Consiglio, del 14 dicembre 2022, relativo alla resilienza operativa digitale per il settore finanziario (DORA).

Proposta di direttiva del Parlamento europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale), COM(2022) 496 final, 28 settembre 2022 (ritirata).

Regolamento (UE) 2023/1113 del Parlamento europeo e del Consiglio, del 31 maggio 2023, riguardante i dati informativi che accompagnano i trasferimenti di fondi e di determinate cripto-attività.

Regolamento (UE) 2023/1114 del Parlamento europeo e del Consiglio, del 31 maggio 2023, relativo ai mercati delle cripto-attività (MiCA).

Regolamento (UE) 2024/1620 del Parlamento europeo e del Consiglio, del 31 maggio 2024, che istituisce l'Autorità per la lotta al riciclaggio e al finanziamento del terrorismo (AMLA).

Regolamento (UE) 2024/1624 del Parlamento europeo e del Consiglio, del 31 maggio 2024, relativo alla prevenzione dell'uso del sistema finanziario a fini di riciclaggio o finanziamento del terrorismo.

Direttiva (UE) 2024/1640 del Parlamento europeo e del Consiglio, del 31 maggio 2024, relativa ai meccanismi che gli Stati membri devono istituire per prevenire l'uso del sistema finanziario a fini di riciclaggio o di finanziamento del terrorismo (VI Direttiva antiriciclaggio).

Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale (regolamento sull'intelligenza artificiale).

Direttiva (UE) 2024/2853 del Parlamento europeo e del Consiglio, del 23 ottobre 2024, sulla responsabilità per danno da prodotti difettosi.

Regolamento (UE) 2026/1744 del Parlamento europeo e del Consiglio, dell'8 luglio 2026, che modifica i regolamenti (UE) 2024/1689, (UE) 2018/1139 e (UE) 2023/1230 (omnibus digitale sull'IA).

Normativa nazionale

D.P.R. 31 marzo 1988, n. 148, Approvazione del testo unico delle norme di legge in materia valutaria.

D.L. 3 maggio 1991, n. 143, convertito, con modificazioni, dalla L. 5 luglio 1991, n. 197.

D.Lgs. 20 febbraio 2004, n. 56.

D.Lgs. 21 novembre 2007, n. 231, e successive modificazioni.

D.Lgs. 13 agosto 2010, n. 141.

D.Lgs. 25 maggio 2017, n. 90.

D.Lgs. 4 ottobre 2019, n. 125.

D.M. del Ministero dell'economia e delle finanze 13 gennaio 2022.

Provvedimenti della Banca d'Italia e della UIF

BANCA D'ITALIA, *Disposizioni in materia di organizzazione, procedure e controlli interni volti a prevenire l'utilizzo degli intermediari a fini di riciclaggio e di finanziamento del terrorismo*, 26 marzo 2019, e successive modificazioni.

UIF, *Istruzioni sui dati e le informazioni da inserire nelle segnalazioni di operazioni sospette*, Provvedimento del 4 maggio 2011.

UIF, *Provvedimento recante gli indicatori di anomalia*, 12 maggio 2023.

UIF, *Istruzioni per la rilevazione e la segnalazione delle operazioni sospette*, Provvedimento del 18 dicembre 2025.

Giurisprudenza

CORTE DI GIUSTIZIA DELL'UNIONE EUROPEA, Sez. I, 7 dicembre 2023, causa C-634/21, *OQ c. Land Hessen (SCHUFA Holding – Scoring)*, ECLI:EU:C:2023:957.

Sitografia

arXiv: arxiv.org

Banca d'Italia: www.bancaditalia.it

Centre for European Policy Studies (CEPS): www.ceps.eu

Commissione europea: commission.europa.eu

EUR-Lex: eur-lex.europa.eu

FATF-GAFI: www.fatf-gafi.org

Normattiva: www.normattiva.it

Unità di Informazione Finanziaria per l'Italia: uif.bancaditalia.it

Ultima consultazione dei siti: luglio 2026.

Dichiarazione del ricorso a strumenti di Intelligenza Artificiale (IA)
Il/La sottoscritto/a Riccardo Tanci, matricola numero 889801
iscritto/a prsso Ca' Foscari al corso di Laurea magistrale in Economia e Finanza, autore della tesi di laurea Antiriciclaggio: l'impatto dell'intelligenza artificiale sul monitoraggio delle transazioni relatore
Andrea Minto, consapevole delle sanzioni previste dall'art. 76 del Testo unico, D.P.R. 28/12/2000 n.445, e della decadenza dei benefici prevista dall'art. 75 del medesimo Testo unico in caso di dichiarazioni false o mendaci, sotto la propria personale responsabilità ai sensi degli artt. 46 e 47 del D.P.R. 28/12/2000 n. 445,
DICHIARA
di aver fatto ricorso a tecnologie di intelligenza artificiale generativa per supportare vari aspetti della ricerca e della scrittura del suo elaborato finale/della mia tesi di laurea, in particolare
Utilizzo dello strumento **Claude Opus 5.5 (Anthropic)** per revisione e ottimizzazione dei testi per migliorare chiarezza e coerenza; generazione di grafici e diagrammi per la presentazione dei dati; supporto alla scrittura automatica di frasi; per tutti i capitoli della presente tesi; l'attendibilità delle informazioni è stata verificata personalmente dell'autore della presente tesi affidandosi a fonti autorevoli citate.
DICHIARA INOLTRE
che tutto il materiale generato dall'IA è stato attentamente verificato, rivisto e rielaborato dall'autore/dall'autrice, che si assume la piena responsabilità per la qualità, l'accuratezza e l'integrità scientifica del lavoro.

Riccardo Tanci 889801