Prevenzione e investigazioni: l'uso di IA, big data e soluzioni tecnologiche in ambito finanziario e nel contrasto al riciclaggio (AML) e al finanziamento del terrorismo (CFT)*[1].

di Fabio Di Vizio[2]

Sommario: - 1. Introduzione. - 2. Fintech, RegTech e SupTech sino all'Intelligenza artificiale. - 3. Il quadro normativo: cenni. - 4. Le nozioni di base. - 4.1. L'intelligenza artificiale: tra miti, realtà e limiti. - 4.1.1. I tentativi definitori. - 4.1.2. Il metodo induttivo e l'approccio deduttivo. - 4.1.3. I limiti reali dei sistemi di IA. – 4.2. Il *machine learning* e i diversi approcci all'apprendimento automatico. - 4.3. Big data, advanced analytics, analisi testuali e AI. - 4.4 Explainable AI e black box. - 5. Le prime applicazioni di IA, big data e ML nelle esperienze di supervisione e di controllo delle autorità pubbliche: la vigilanza bancaria e il contrasto all'evasione fiscale. - 6. Le strategie di contrasto per mezzo dell'IA del riciclaggio e del finanziamento al terrorismo. - 6.1. Le premesse. - 6.2. Le opportunità delle nuove tecnologie nel contrasto del riciclaggio e del finanziamento del terrorismo nell'analisi del GAFI. - 6.3. L'impiego dell'IA nell'AML: le entità regolamentate. - 6.4. L'impiego dell'IA nell'AML: le autorità. – 6.5. L'impiego dell'IA nel contrasto del terrorismo. - 6.5.1. Il momento definitorio, le tecniche di finanziamento e di contrasto del terrorismo. - 6.5.2. La lotta al terrorismo internazionale, tra disciplina di contrasto della criminalità organizzata e del riciclaggio: dal fenomeno finanziario a quello ideologico. - 6.5.3. Nuove fenomenologie di terrorismo internazionale: impiego di applicazioni IA. - 6.5.4. Il contrasto dell'IA al terrorismo. - 6.5.5. In particolare, il contrasto dell'IA alla radicalizzazione terroristica.

-1. Introduzione.

Come annuncia la Risoluzione del Parlamento europeo del 3 maggio 2022 sull'intelligenza artificiale in un'era digitale[3] «il mondo è **sull'orlo della quarta rivoluzione industriale**». Se le prime tre manifestazioni di quest'ultima sono state avviate dall'introduzione del vapore, dell'elettricità e dei computer, la quarta trae energia da un'abbondanza di **dati** unita a potenti **algoritmi** ed al rafforzamento della capacità di **calcolo.** È questa la **nuova sostanza** dalla rivoluzione digitale, caratterizzata da portata globale, rapida convergenza ed enorme impatto delle scoperte tecnologiche su Stati, economie, società, relazioni internazionali e ambiente. La **transizione digitale** realizza un cambiamento radicale di paradigmi che incide in modo diverso sulla vita di ciascun individuo e di varie parti della società, a seconda della diversità di obiettivi, posizioni geografica e contesto socioeconomico. Il rispetto dei diritti fondamentali e del principio di non discriminazione richiede declinazioni innovative, espressive di una nuova e consapevole cittadinanza globale conservando fermo il rapporto di funzionalizzazione delle tecnologie all'uomo. In un quadro già attuale di globale competitività digitale, enorme è il valore economico delle capacità e delle risorse impegnate nella ricerca, nello sviluppo e nella commercializzazione delle applicazioni di intelligenza artificiale, quale complesso di tecnologie emergenti in grado di influenzare il potere geopolitico di interi paesi.

Big data*, intelligenza artificiale, *machine learning*, *cloud* e *distributed ledger technology sono le nuove tecnologie dell'innovazione digitale che dal mondo finanziario si è trasferita alla realtà sociale pervadendola in profondità e ponendo **nuove sfide** a regolatori politici e autorità investigative per i possibili utilizzi criminali di esse. Come è stato osservato «l'innovazione tecnologica ha determinato l'emersione di nuovi strumenti e operatività, ha ampliato la platea degli operatori cui si applicano gli **obblighi antiriciclaggio** e fatto emergere **nuovi rischi** difficilmente presidiabili con i metodi tradizionali. Occorre attrezzarsi per fronteggiarli, dal punto di vista sia **regolamentare** sia dell'adeguamento delle **metodologie** e delle **risorse** impiegate da parte delle Istituzioni. Il ricorso

[1] Relazione al corso "*La digital transformation: evoluzione del contesto e profili di impatto di diritto penale sostanziale e processuale*", Corso organizzato dalla Scuola Superiore della Magistratura in collaborazione con la Scuola di Polizia economico- finanziaria della Guardia di finanza, Cod. FPFP23015, 16 novembre 2023, Lido di Ostia (Roma).

[2] Sostituto Procuratore presso la Direzione Distrettuale Antimafia della Procura di Firenze.

[3] *European Parliament resolution of 3 May 2022 on artificial intelligence in a digital age* (2020/2266(INI)).

all'intelligenza artificiale e alle tecnologie avanzate di trattamento delle informazioni è una **necessità** imprescindibile per entrambe le componenti del sistema di prevenzione, pubblica e privata, e può migliorarne i processi di lavoro. […] A meno che il futuro non ci riservi scenari inaspettati, ancora per molti anni la **variabile umana** sarà fondamentale per guidare e utilizzare al meglio le nuove tecnologie; altrimenti, lasciando le macchine a sé stesse, si corre il forte rischio di ridurre l'efficacia del sistema di prevenzione»[4].

Occorre aver chiare la **problematicità** e l'**inevitabilità** del percorso che si prospetta dinanzi. Le diverse declinazioni dell'IA impongono di orientare la scelta verso un **approccio responsabile** alle innovazioni che essa comporta ed origina.

Il **rapporto tra uomo e macchina cambia**, così come mutano le attese del primo rispetto alla seconda, non richiesta più solo di eseguire comandi secondo programmi e percorsi strettamente predeterminati dal primo ma di offrire "nuova conoscenza" grazie a originali capacità computazionali e di perfezionamento (apprendimento) con l'apprendimento dei dati tratti dall'esperienza, originando **nuovi modelli di sapere**. Si tratta del **cuore della questione**: l'uomo non si fida più della macchina perché infallibile nel replicare meccanicisticamente la conoscenza del primo, ma si affida, in parte, alla stessa accettandone un **margine di fallibilità** quale prezzo del progresso e della scoperta di sapere estraneo a chi ha impostato il sistema. L'evoluzione è figlia di questa **fiducia** dell'uomo delle potenzialità di perfezionamento (apprendimento) della "machina", base della crescita cognitiva di quest'ultima e alla quale va concesso di rivendicare il brocardo latino "**errando discitur**". L'irresponsabilità maggiore sarebbe dimenticare la realtà di questa accettazione del dubbio e della possibilità dell'errore quale mezzo inevitabile per acquisire nuova conoscenza.

Venendo alle **applicazioni dell'IA** in funzione di **controllo e vigilanza** pubblica, in molti Paesi si stanno affermando sistemi in grado di prevedere, sorvegliare, identificare in maniera rapida ed efficace **comportamenti illeciti** o scorretti dei cittadini. In tali casi, l'analisi algoritmica si colloca in una **fase preistruttoria, con una funzione predittiva** volta ad indirizzare l'attività di vigilanza e di controllo esercitata dall'amministrazione, consentendo l'elaborazione di dati che sfuggono all'analisi umana per divenire un elemento di supporto alla decisione adottata dal funzionario pubblico[5].

Si pensi alla **sorveglianza automatica** delle reti di comunicazione che consente di individuare **rischi terroristici** non rilevabili da un analista di *intelligence* o agli algoritmi **fiscali** capaci di combinare una pluralità di dati di diversa provenienza al fine di tracciare un profilo del contribuente e rilevare anomalie che sfuggono ai tradizionali controlli o ancora al c.d. *digital welfare stat*e per la gestione e sorveglianza dell'**assistenza sociale**.

Tale attività sovente si pone di frizione con le **garanzie di trasparenza o di legalità algoritmica**, già in parte sviluppate con riferimento a provvedimenti automatizzati[6]. Gli algoritmi più complessi, infatti, generano risposte non sempre riproducibili, grazie ad un'analisi della realtà superiore alle capacità umane e allo sfruttamento di *big data* o *data set* molto ampi. Il **fatto "accertato" deriva,** perciò, da un **insieme di dati decontestualizzati** che trovano **nell'algoritmo un nuovo e autonomo significato**[7]. I più moderni e potenti algoritmi predittivi partono dall'analisi di

[4] E. SERATA, *Innovazione tecnologica e prevenzione del riciclaggio*, testo (inedito) della relazione tenuta al convegno "Regolazione e Supervisione finanziaria nell'era dirigitale", Unitelma, 19 maggio 2023.

[5] F. LAGIOIA - G. SARTOR, *Profilazione e decisione algoritmica: dal mercato alla sfera pubblica*, in *federalismi.it*, 2020, 11, 85 ss.; F. COSTANTINO, *Rischi e opportunità del ricorso amministrativo alle predizioni dei big data*, in *Dir. pubbl.*, 2019, I, 43 ss.

[6] L'algoritmo di mobilità del personale della scuola ha sollevato le prime questioni di trasparenza e conoscibilità degli algoritmi, cfr. Cons. Stato, Sez. VI, 8 aprile 2019, n. 2270; Cons. Stato 13 dicembre 2019, n. 8472; Cons. Stato 4 febbraio 2020, n. 881. Per commenti cfr. E. CARLONI, *I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo*, in *Dir.amm.*, 2020, 271 ss.; D.U. GALETTA, *Algoritmi, procedimento amministrativo e garanzie: brevi riflessioni, anche alla luce degli ultimi arresti giurisprudenziali in materia*, in *Riv. it. dir. pubbl. comunitario*, 2020, 3-4, 501 ss.; S. SASSI, *Gli algoritmi nelle decisioni pubbliche tra trasparenza e responsabilità*, in *Analisi Giuridica dell'economia*, 2019, spec. 117-118.

[7] G. AVANZINI, *Intelligenza artificiale e nuovi modelli di vigilanza pubblica in Francia e Olanda*, in *Giornale Dir. Amm.*,

dati spesso destrutturati ed effettuano la profilazione o l'indicizzazione di comportamenti umani che vengono ad integrare i presupposti per l'esercizio del potere pubblico. Per tale ragione le nuove forme di vigilanza operano grazie ad una previa e generalizzata raccolta di **informazioni apparentemente scollegate** e analizzabili solo attraverso programmi informatici che hanno una autentica portata **creativa della realtà fenomenica**; analizzando dati anche estranei e non direttamente attinenti all'esercizio del potere, tali sistemi amplificano il pericolo di compressione o di violazione delle posizioni soggettive dei cittadini, proiettando gli effetti della profilazione fuori del perimetro dell'azione amministrativa, prima e indipendentemente dalla decisione di impulso procedimentale. Il **modello di controllo e di vigilanza tradizionale si trasforma**, allontanandosi da una **logica mirata** (che opera in presenza di un dubbio o di un sospetto preesistente che in qualche modo disvela la condotta illecita da accertare in modo più approfondito), o su **verifiche a campione**, imponendo di riflettere con i problemi posti dai nuovi sviluppi. Il sistema può costituire in sé una ***black box***[8], incomprensibile per gli stessi programmatori; la **disciplina** di questa fase, allo stato, è **deficitaria**, nonostante la rilevanza degli effetti che derivano sul successivo esercizio del potere e sulle posizioni soggettive dei cittadini incidentalmente coinvolti dall'analisi algoritmica.

L'esigenza di **nuove forme di garanzie nel caso di decisioni in tutto o in parte automatizzate** è riconosciuta da tempo a livello a livello internazionale, dove di rimarca la necessità di incentivare un'**Intelligenza Artificiale antropocentrica**[9], creata per l'uomo e a supporto delle esigenze umane e come tale rispettosa delle libertà fondamentali del cittadino che si esprimono anche nei confronti della pubblica amministrazione digitale anche laddove essa utilizzi l'analisi predittiva.

La riflessione che segue muove dalla necessità di ambientare convenientemente l'analisi dei nuovi strumenti preventivi ed investigativi nel contesto di una **mutata realtà** che prima che dimensione tecnologica ha assunto nitida natura economico e sociale (§ 2) ricostruendo il **quadro normativo** in via di evoluzione (§ 3); di essa occorre che tengano conto le **autorità pubbliche**, preposte alla tutela di interessi pubblici, siano esse amministrative o giudiziarie, trasformando le declinazioni con cui esse hanno sinora esercitato le loro prerogative tradizionali. Si delinea un sistema che registra l'emersione del **bisogno di nuove garanzie individuali**, il cui soddisfacimento è in parte consistente affidata a **privati** e al **diritto giurisprudenziale**, tenuto conto della prudenza dei **regolatori politici**, nel contribuire anche a chiarire le nozioni di base e le definizioni comuni dalle quali dovrebbe muoversi (§4). Si verranno, così, esaminando le **prime applicazioni di IA, big data e ML nelle esperienze di supervisione e di controllo** delle autorità pubbliche: dalla **vigilanza bancaria al contrasto all'evasione fiscale** (§5), per diffondersi sulle strategie di contrasto per mezzo dell'IA del **riciclaggio** e del **finanziamento** al terrorismo (§6), tenuto conto dei profili comuni ma anche delle consistenti particolarità dei due fenomeni criminali.

- 2. Fintech, RegTech e SupTech sino all'Intelligenza artificiale.

L'applicazione delle nuove tecnologie alla finanza va sotto il nome di ***Fintech***. Il fenomeno si sostanzia nell'offerta di attività e servizi finanziari in forme originali, con nuovi prodotti, emessi sotto forma di *token*, nel contesto di un ecosistema popolato da nuovi soggetti (dalle *start-up* ai colossi

2022, 3, 316, osserva: «Alla base del loro sfruttamento vi è l'idea che in questa massa di dati vi sia un valore cognitivo intrinseco e specifico, che manca in universi più piccoli. Il processo di analisi mira quindi a far parlare il dato attraverso innumerevoli e disparate correlazioni, che non sono lineari né tanto meno logiche e che non hanno lo scopo di spiegare la realtà ma solo di pervenire ad una soluzione».

[8] G. LO SAPIO, *La* black box*: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, in *Federalismi.it*, 16, 2021; F. PASQUALE, *The Black Box Society. The Secret Algorithms that Control Money and Information*, Cambridge, London, 2016.

[9] In questo senso cfr. la comunicazione 25 aprile 2018, Strategia per l'I.A. (COM (2018) 237 final) e le Linee *Guida etiche "Creare fiducia nell'intelligenza artificiale antropocentrica"* 8 aprile 2019 (COM (2020) 65 final).

tecnologici Amazon e Google) che affiancano gli operatori tradizionali (banche e imprese di investimento)[10]. Il *Fintech* apporta benefici all'industria finanziaria: la concorrenza, la velocità delle attività e dei servizi e l'inclusione finanziaria vengono incrementate a fronte di una riduzione dei costi dovuta alla disintermediazione[11]. Tale sviluppo non è privo di rischi, legati a malfunzionamenti nella fornitura di tecnologia sottesa ai servizi finanziari, al problematico rispetto delle normative pertinenti nonché all'opacità del regime applicabile, per l'assenza di una definizione generalmente condivisa e la prudenza nell'adozione di approcci regolamentari, nonostante la supervisione degli attori sistemici del *Fintech* sia sempre più importante per garantire la stabilità del sistema finanziario e proteggere i diritti degli utenti[12].

L'espressione *Fintech*[13] appare riferibile ad un fenomeno che può essere ricostruito sotto due aspetti: da un lato, l'affermazione sul mercato di nuovi modelli di *business*, connotati dall'offerta di uno o più servizi o prodotti finanziari in modalità automatizzata o comunque innovativa; dall'altro, l'impiego di tecnologie emergenti che stanno ridisegnando l'industria finanziaria - quali, appunto, *machine learning (ML)*, intelligenza artificiale (AI), *distributed ledger technology* (DLT) - utilizzabili sia da nuovi soggetti, sia da intermediari tradizionali. Vengono in mente, esemplificativamente, il servizio di consulenza in materia di investimenti o di gestione del portafoglio, offerti con modalità automatizzate tramite un algoritmo, o l'offerta sul mercato di criptoattività, tramite l'utilizzo della *distributed ledger technology* (DLT)[14].

---

[10] Come osserva B. BARMANN, Fintech: *primi tentativi di regolazione*, in *Giornale Dir. Amm., 2021, 6, 811*, l'effetto dirompente del *Fintech* si manifesta non soltanto su un piano oggettivo, relativo al tipo di prodotto o servizio, ma anche su quello soggettivo dei partecipanti al mercato. Accanto ai tradizionali intermediari bancari e finanziari (c.d. *incumbent)* sono presenti sul mercato nuovi soggetti: *start-up* innovative che prestano servizi finanziari digitali ovvero offrono soluzioni tecnologiche in ambito finanziario, nonché altre imprese che forniscono supporto tecnologico nell'offerta di servizi finanziari *(technology providers* e *ICT companies).* Nell'ambito di quest'ultima categoria rientrano anche i colossi tecnologici come Apple, Google e Facebook, che sempre più hanno esteso la loro operatività nel settore finanziario (si pensi agli accordi di Apple con Goldman Sachs e di Google con altre banche americane o ad Ant Group e Alipay nate come costole di Alibaba).

[11] L'innovazione digitale apporta notevoli benefici e significative opportunità al settore finanziario. L'ingresso di nuovi attori accresce la concorrenza e spinge gli operatori tradizionali verso l'innovazione. L'impiego delle nuove tecnologie comporta, inoltre, una disintermediazione senza precedenti: la consulenza automatizzata avviene direttamente con il cliente, senza che sia necessario l'intervento umano (a seconda dei casi, può essere previsto un coinvolgimento limitato ad alcune fasi del relativo processo); le piattaforme P2P *lending* (peer-to-peer), così come quelle di *crowdfunding*, mettono direttamente in contatto il soggetto finanziatore/investitore con il soggetto che necessita di fondi. L'impiego dell'IA consente, inoltre, di processare a velocità senza precedenti enormi quantità di dati e informazioni e di offrire servizi su misura ai clienti. In generale, si registra un miglioramento della qualità e della velocità dei servizi, nonché una riduzione dei costi con una conseguente maggiore inclusività del settore finanziario.

[12] M. DOBLER e altri, *E-Money: Prudential Supervision, Oversight, and User Protection,* in *Monetary and Capital Markets and Legal Departments Discussion Paper 21/027*, International Monetary Fund, Washington, DC, 2021. Secondo report in materia pubblicati da KPMG, il mercato degli investimenti globali nel Fintech ha raggiunto quota 98 miliardi di dollari nel primo trimestre del 2021. Tra i rischi legati alla diffusione delle nuove tecnologie nel settore finanziario, alcuni sono comuni ad altri settori dell'economia, come la *privacy*, la sicurezza dei dati o la *price optimisation*, con il rischio di adottare comportamenti sfavorevoli nei confronti della clientela. Vi è, poi, la mancanza di chiarezza riguardo al regime normativo applicabile, con difficoltà ad estendere le regole esistenti a servizi e attività finanziari innovativi. Può accadere, difatti, che specifiche attività, servizi e prodotti non ricadano nelle categorie tradizionali perché prestate con modalità innovative per il settore, ma aventi carattere "abilitante". Si pensi ai *cypto-assets* che, a seconda delle caratteristiche che presentano, possono o meno essere ricondotti nella categoria dei prodotti finanziari, con conseguente applicazione della disciplina europea in materia (Mifid, Csdr, Mar, Regolamento Prospetto). Si considerino, inoltre, le imprese che forniscono servizi ICT agli intermediari finanziari (in *outsourcing, partnership*, ecc.).

[13] Secondo il *Financial Stability Board*, il *Fintech* può essere definito come "*technologically enabled innovation in financial services that could result in new business models, applications, processes or products with an associated material effect on financial markets and institutions and the provision of financial services*" (*Financial Stability Board*, in *Basel, Switzerland, definizione aggiornata* al 5.5.2022, reperibile in *https://www.fsb.org/work-of-the-fsb/financial-innovation-and-structural-change/fintech*). Il *Regulatory Technology* (RegTech) è un sottoinsieme di FinTech che utilizza le nuove tecnologie per conformarsi ai requisiti normativi in modo più efficiente ed efficace rispetto alle capacità esistenti.

[14] Anche il GAFI rileva come l'espressione FinTech si riferisca in generale all'uso di tecnologie digitali nuove ed emergenti nel settore finanziario per un'ampia varietà di scopi. Inizialmente, "FinTech" si riferiva principalmente

A fronte della complessità di tali tecnologie innovative la **semplificazione che esse consentono nell'accesso alla realtà finanziaria**, favorita dalla disintermediazione di molte fasi dello stesso, è sotto gli occhi tutti, anche se talvolta resa quasi inconsapevole dall'estrema rapidità dell'evoluzione e dei cambiamenti. La rete internet e l'utilizzo di *smartphones* hanno reso immediato l'utilizzo dei servizi finanziari, permettendo, ad esempio, di effettuare *trading online* o pagamenti direttamente tramite applicazioni sul telefono[15]. Emergono **nuove tecnologie, nuovi operatori, prodotti, servizi, canali distributivi digitali e *cross-border*** diversi da quelli tradizionali, con modalità di accesso ai prodotti e ai servizi senza alcuna intermediazione da parte di operatori vigilati. Si affermano **nuovi modelli di *business***, si diffondono strumenti di pagamento e di investimento alternativi, che espongono a rischi anche di riciclaggio e di altri utilizzi criminali e creano difficoltà nell'applicazione dei tradizionali presidi del sistema di prevenzione. *Distributed Ledger Technologies*, *Crypto-assets*, *Smart Contracts*, *Centralized vs. Decentralized Finance (CeFi vs. DeFi)*, *Instant Payments*[16], l'emissione e la circolazione di strumenti finanziari in forma digitale disciplinate dal c.d. DL *Fintech*[17].

Se è impossibile arginare lo **sviluppo tecnologico**, occorre favorire la digitalizzazione senza ridurre le **difese contro la criminalità**, trovando un **punto di equilibrio tra rischi e opportunità**, adeguando le regole per evitare arbitraggi normativi e assicurare controlli efficaci sulle attività *online*, utilizzando l'innovazione tecnologica per migliorare le attività di supervisione e di analisi.

In ragione di ciò, la **sfida che si pone davanti ai regolatori pubblici** è ardua: da un lato, considerate le potenzialità del *Fintech*, è chiara la necessità di agevolarne quanto più possibile lo sviluppo; allo stesso tempo, tuttavia, è necessario non sottovalutare i rischi che i nuovi soggetti e gli innovativi prodotti e servizi offerti possono comportare in termini di tutela degli investitori/risparmiatori nonché di stabilità del sistema finanziario. Le cripto-attività, le nuove forme di finanza digitale decentralizzata, unite alla rarefazione del rapporto personale tra intermediari e clienti, richiedono la capacità delle Istituzioni di fronteggiare le nuove problematiche connesse con l'utilizzo delle nuove tecnologie.

In tale contesto, l'**intelligenza artificiale** (d'ora in poi anche IA o AI, acronimo di *Artificial Intelligence*) rappresenta **una delle tecnologie strategiche del XXI secolo**, alle quali i decisori politici riconnettono importanti benefici in termini di efficienza, precisione e comodità, con contributo positivo all'economia europea. Le attuali applicazioni di AI hanno migliorato, tra l'altro, le cure sanitarie, accresciuto l'efficienza dell'agricoltura, contribuito alla mitigazione e all'adattamento ai cambiamenti climatici nonché migliorato l'efficienza della produzione.

Anche l'impatto dei **Big Data,** degli algoritmi di *Machine Learning* e in generale dell'Intelligenza Artificiale sulle **scienze sociali** è stato dirompente negli ultimi anni, grazie allo

---

all'applicazione di innovazioni basate sulla tecnologia per fornire nuovi prodotti e servizi finanziari rivolti al cliente (ad esempio, soluzioni di pagamento mobile, prestiti sui mercati online, strumenti algoritmici di risparmio e investimento, pagamenti in valuta virtuale, raccolta di capitali e di depositi). FinTech ora comprende anche l'uso di nuove ed emergenti tecnologie per fornire funzioni aziendali automatizzate di *mid-office* e *back-office*, come l'uso di algoritmi, *big data*, IA, apprendimento automatico e analisi dei collegamenti per la liquidazione, il regolamento, l'intermediazione e altre attività, ad esempio, relative a titoli finanziari, derivati e pagamenti, nonché per attività di adeguamento normativo.

[15] Ne costituiscono esempi Alipay, tramite la quale è possibile effettuare e ricevere pagamenti attraverso lo *smartphone*, M-Pesa in Kenya o WeChat in Cina. Non è un caso che un forte sviluppo di servizi digitalizzati sia diffuso in molti Paesi in via di sviluppo, con percentuali elevate di popolazione *unbanked*.

[16] Cfr. Newsletter dell'UIF n. 5 del 2022, Aggiornamenti in materia di *virtual asset*; FATF, *Virtual Assets: Targeted Update on Implementation of the FATF Standards on Virtual Assets and Virtual Asset Service Providers*, giugno 2023; Banca d'Italia, Consultazione pubblica su documento di lavoro relativo alla prima fase dell'attività di ricerca sugli *smart contract*, giugno 2023; Negoziato europeo sulla proposta legislativa della Commissione europea per sostenere l'uso dei pagamenti istantanei.

[17] DL 17 marzo 2023, n. 25, come modificato dalla legge di conversione 10 maggio 2023, n. 52.

sviluppo di computer sempre più potenti, alla diffusione di dati sempre più granulari su vari aspetti della vita quotidiana degli individui e allo sviluppo di algoritmi sempre più efficienti[18].

***Big data analytics* e IA** hanno meritato, di recente, molta attenzione nel **dibattito pubblico**, anche con riferimento al loro impiego per prevenire e individuare fenomeni criminali.

Nello specifico ambito dell'antiriciclaggio, si discute molto, anche in Italia, dell'utilizzo della tecnologia per gestire i rischi e gli obblighi normativi (il cd. **RegTech**), in correlazione con la rivoluzione della tecnologia applicata alla finanza (il FinTech)[19]. Il punto nodale è quello di verificare la possibilità che, tramite questi strumenti 'intelligenti', si possa migliorare l'analisi predittiva dell'ingente volume di dati su transazioni e clienti e individuare in maniera più accurata schemi di illeciti e situazioni ad alto rischio non censite dai più tradizionali modelli basati su regole deterministiche. Oltrepassando l'attività AML/CFT affidata a controlli di natura documentale e manuale e l'utilizzo di modelli **e strumenti AML tradizionali (**con soluzioni basate sull'utilizzo di **'motori inferenziali' e** di regole deterministiche)[20], l'impiego di strumenti evoluti consentirebbe di andare oltre e rivedere il ventaglio di indicatori di anomalia abitualmente utilizzati in ambito AML/CFT. Come risultato, sia le autorità di supervisione AML, da un lato, che i soggetti obbligati, dall'altro, hanno cominciato ad adottare – o a pianificare di farlo – queste **soluzioni tecnologiche evolute nella loro attività operativ**a. In particolare, le autorità AML e le unità di intelligence finanziarie (FIU) impiegano oggi i modelli basati su intelligenza artificiale, tra le altre cose, come **supporto nell'analisi delle segnalazioni di operazioni sospette** ricevute e per orientare in maniera più efficace la successiva attività di ispezione e vigilanza[21]. Sull'altro fronte, banche, assicurazioni, istituti di pagamento e gestori di giochi/scommesse hanno cominciato ad investire ingenti risorse nell'acquisizione e impiego di tecnologie, software e risorse umane con capacità analitiche, da dedicare all'uso di soluzioni avanzate nell'attività AML/CFT. Tuttavia rimangono dubbi sull'effettivo stato di utilizzo di questi strumenti da parte dei soggetti obbligati e sulle problematiche incontrate nella loro adozione.

L'AI offre grandi opportunità anche nel settore delle attività di **prevenzione e di contrasto dei reati e della giustizia penale**, migliorando i **metodi di lavoro** delle autorità investigative e giudiziarie e consentendo, in particolare, di lottare in modo **più efficace** contro alcuni tipi di reati, in particolare nel settore dei reati finanziari, del riciclaggio di denaro e del finanziamento del terrorismo, oltre che contro alcuni tipi di reati informatici. Se l'aumento esponenziale di informazioni di natura digitale, strutturate e non-strutturate, a disposizione dei soggetti obbligati può fornire opportunità enormi per migliorare l'identificazione delle condotte criminali nell'ambito del AML/CFT, laddove li indicatori di anomalia tradizionali e i cd. "motori a regole" finora adottati non colgono tutti gli schemi illeciti emergenti, sono elevati anche i rischi derivanti dall'adozione di questi strumenti evoluti, soprattutto se non gestiti da risorse umane competenti e capaci di leggere, dietro un'anomalia statistica rilevata da una "macchina intelligente", un potenziale comportamento criminale. Il presente contributo è dedicato a questo particolare impiego delle applicazioni dell'AI che include, ad esempio, le tecnologie di riconoscimento facciale, il riconoscimento automatizzato delle targhe,

---

[18] J. MARCUCCI, *Big Data, Machine Learning e Artificial Intelligence nell'analisi economica e statistica,* in F. FEDERICO, J. MARCUCCI, M. BEVILACQUA, DJ MARCHETTI, *Rapporto 2/2021 – L'impiego dell'intelligenza artificiale nell'attività di Banca d'Italia,* in *BioLaw*, 24 dicembre 2021.

[19] In tema cfr.: *FATF position on FinTech and RegTech* (2020); il *white paper* di BIS (*Bank of International Settlements*) e di UIF '*Suptech applications for anti-money laundering*' (COELHO, DE SIMONI, PRENIO 2019); il report della *European Banking Authority* (EBA) "*EBA Report on Big Data and Advanced Analytics*" (2020); il position paper della *European Banking Federation* (EBF) "*EBF position paper on AI in the banking industry*" (2019); il report del *Financial Stability Board* (FSB) "*Artificial intelligence and machine learning in financial services: Market developments and financial stability implications*" (2017).

[20] Si tratta dei cosiddetti **'motori a regole'** (o motori inferenziali) soluzioni avanzate che, in ambito AML/ CFT, fanno riferimento a modelli che generano un determinato *output* al raggiungimento/superamento di soglie predefinite per le diverse variabili incluse nei modelli stessi. Occorre rimarcare che l'utilizzo di queste soluzioni è tuttora predominante.

[21] COELHO, DE SIMONI, PRENIO, *Suptech applications for anti-money laundering (testo in inglese)pdf* , 9 dicembre 2019.

l'identificazione vocale, tecnologie di lettura labiale, analisi di segnali acustici (algoritmi di rilevamento di colpi di arma da fuoco), ricerca autonoma e analisi di database identificati, previsioni (polizia predittiva e analisi della scena del crimine), strumenti di rilevamento dei comportamenti, strumenti autonomi per identificare le frodi finanziarie e il finanziamento del terrorismo, monitoraggio dei social media (estrazione e raccolta di dati per l'estrazione di connessioni), numeri IMSI e sistemi di sorveglianza automatica che integrano diverse capacità di rilevamento (come il rilevamento cardiaco e le videocamere termiche). In ambito giudiziario, ancora, gli strumenti di AI possono essere utilizzati nel calcolo delle probabilità di recidiva e nelle decisioni di sospensione condizionale o di condanna.

Nonostante i benefici che essa apporta, l'AI implica, nel contempo, anche una serie di rischi potenziali, quali **processi decisionali opachi, vari tipi di discriminazione**, intrusione nella vita privata, rischi per la protezione dei dati personali, per la dignità umana e la libertà di espressione e informazione. Tali rischi sono ancora p**iù gravi nel settore delle attività di contrasto e della giustizia penale**, in quanto possono incidere sulla presunzione di innocenza, sui diritti fondamentali per la libertà e la sicurezza dell'individuo e su un ricorso effettivo e un processo equo[22].

La presente relazione intende affrontare le potenzialità e le questioni sollevate dall'uso dell'IA in materia preventiva e penale da parte delle autorità pubbliche, di *intelligence* finanziaria, di polizia o giudiziarie. Pur prendendo atto delle opportunità e dei vantaggi prevedibili offerti dall'AI, non vanno dimenticati anche i rischi significativi che essa può comportare ed i limiti che si vengono profilando via via che progredisce il processo normativo e regolatorio.

Le potenzialità dell'AI vanno apprezzate entro i confini della necessità di rispettare appieno i **diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea**, la legislazione dell'Unione in materia di tutela della vita privata e protezione dei dati, in particolare la direttiva (UE) 2016/680 ("direttiva polizia") e la necessità di rispettare **diversi principi fondamentali nel ciclo di vita dell'IA, tra cui la spiegabilità e la trasparenza degli algoritmi**, la tracciabilità, l'esecuzione di valutazioni di impatto obbligatorie sui diritti fondamentali prima dell'attuazione o della diffusione di qualsiasi sistema di IA e audit obbligatori. Tutti questi requisiti sono necessari non solo per garantire la legittimità dei sistemi di IA, ma anche per ottenere la fiducia delle persone per quanto riguarda l'utilizzo di tali sistemi da parte delle autorità di contrasto e delle autorità giudiziarie. Se talune applicazioni di IA hanno raggiunto livelli prestazionali analoghi a quelli di esperti umani e professionisti nell'esecuzione di taluni compiti specifici (per es. le tecnologie applicate al contesto giuridico) e possono offrire risultati con una velocità e una scala notevolmente superiori, resta necessario assicurare che i diritti e le libertà fondamentali sanciti nella Carta siano rispettati per l'intera durata del ciclo di vita dell'IA e delle tecnologie correlate, in particolare durante la loro progettazione, sviluppo, diffusione e impiego, e applicarsi alle attività di contrasto in ogni circostanza, i sistemi di IA dovendo essere concepiti per la protezione e il vantaggio di tutti i membri della società (tenendo conto, nella loro progettazione, delle popolazioni vulnerabili ed emarginate), essere non discriminatori, sicuri, conducendo a decisioni spiegabili e trasparenti, rispettando l'autonomia umana e i diritti fondamentali per poter essere considerati affidabili, in modo da mettere le persone al centro, ponendosi sempre al servizio dell'essere umano e i sistemi di AI dovendo essere progettati in modo da poter sempre essere spenti da un operatore umano.

---

[22] In tema cfr. P. VITANOV, *Relazione sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*, A9-0232/2021, Commissione per le libertà civili, la giustizia e gli affari interni del Parlamento europeo, 13 luglio 2021. Il relatore ha chiesto una moratoria sulla diffusione dei sistemi di riconoscimento facciale a fini di contrasto in quanto l'attuale stato di avanzamento di tali tecnologie e il loro impatto significativo sui diritti fondamentali richiedono un dibattito sociale aperto e approfondito, al fine di esaminare le diverse problematiche sollevate e la giustificazione di una loro diffusione.

In ogni caso, occorre muovere da un **dato di partenza indispensabile** per comprendere l'intero assetto, attuale e prospettico. Come ricorda la relazione Vitanov «nonostante i continui progressi compiuti in termini di velocità di elaborazione e capacità di memoria dei computer, attualmente **non esistono ancora programmi che possano assicurare una flessibilità analoga a quella dell'essere umano** in relazione a domini più ampi o a compiti che richiedono la comprensione di un contesto o un'analisi critica».

- **3. Il quadro normativo: cenni**.

Il 30 ottobre 2023 i leader del **G7** hanno sottoscritto una dichiarazione sul processo di Hiroshima sullo sviluppo dell'AI sottolineando le opportunità innovative e trasformative ricollegate ai sistemi avanzati, con particolare riferimento all'AI generativa[23]. Nella dichiarazione compare esplicitamente il riconoscimento della necessità di gestire i rischi e di proteggere gli individui, la società e i principi condivisi tra cui lo stato di diritto e i valori democratici, mantenendo una visione antropocentrica, con una governance inclusiva. I leader hanno accolto con favore i Principi guida internazionali per le organizzazioni che sviluppano sistemi avanzati di AI, secondo il processo di sviluppo condiviso ad Hiroshima, che prevede un quadro politico globale, comprensivo della cooperazione basata su progetti in collaborazione con la *Global Partnership for Artificial Intelligence* (GPAI)[24] e l'Organizzazione per Cooperazione e Sviluppo Economico (OCSE)[25] e per condurre

[23] L'intelligenza artificiale generativa (o GenAI) è un tipo di intelligenza artificiale che è in grado di generare testo, immagini, video, musica o altri media in risposta a delle richieste (*prompt*). I sistemi di intelligenza artificiale generativa utilizzano modelli generativi, modelli statistici di distribuzione congiunta di una variabile osservabile e di una variabile dipendente (che nel contesto del *data mining* è detta variabile *target*). Un esempio sono i modelli linguistici di grandi dimensioni *(Large Language Model*) che producono dati a partire da un *dataset* di addestramento utilizzato per crearli. Tra i sistemi di intelligenza artificiale generativa degni di nota si ricordano ChatGPT, una chatbot creata da OpenAI utilizzando i modelli linguistici GPT-3 e GPT-4.; altri sistemi includono Bard di Google (basato sul modello LaMDA), Bedrock di Amazon, Ernie Bit di Baidu, Pangu-Σ di Huawei, Claude di Anthropic, Jais in lingua araba e Poe di Quora. Dolly 2.0 è il primo LLM interamente open source e libero da restrizioni anche per finalità commerciali e di ricerca, creato da Databricks. Esistono, inoltre, sistemi capaci di generare immagini 3D come Stable Diffusion, Midjourney e DALL-E. L'intelligenza artificiale generativa ha potenziali applicazioni in una vasta gamma di settori, tra cui lo sviluppo *software*, il *marketing* e la moda, l'editoria, la predizione di struttura proteica e la scoperta di farmaci (a partire da catene di aminoacidi o rappresentazioni di molecole). Gli investimenti nell'IA generativa sono aumentati nei primi anni 2020: Microsoft ha investito 10 miliardi di dollari in OpenAI, Google e Baidu e numerose aziende più piccole che sviluppano modelli di IA generativa.

[24] La *Global Partnership on Artificial Intelligence* è un'iniziativa internazionale, annunciata dal Primo ministro del Canada e dal Presidente della Repubblica francese prima del vertice del G7 2018, in attuazione della dichiarazione Canada-Francia su AI del giugno 2018, alla quale hanno preso parte Australia, Canada, Francia, Germania, Giappone, India, Italia, Messico, Nuova Zelanda, Regno Unito, Repubblica di Corea, Singapore, Slovenia, Stati Uniti d'America e Unione Europea. Dopo aver annunciato il mandato per la costituzione del gruppo internazionale sull'intelligenza artificiale durante la conferenza *multistakeholder* G7 sull'intelligenza artificiale nel dicembre 2018, in occasione del vertice G7 nell'agosto 2019 a Biarritz, i Capi di Stato e di Governo hanno riconosciuto la GPAI proposta da Canada e Francia nell'ambito della strategia di Biarritz per una trasformazione digitale aperta, libera e sicura, tenuto conto delle linee guida indicate dalla *Recommendation of the Council on Artificial Intelligence* approvata dall'OCSE nel 2019 per affermare il riconoscimento di principi costituenti standard generali OCSE che mirano ad assicurare la tutela della *privacy* e della sicurezza digitale nelle politiche di implementazione dell'intelligenza artificiale.

[25] La ricordata Raccomandazione Ocse, sebbene priva di valore giuridicamente vincolante, individua specifici **principi complementari su cui si basa la gestione responsabile di una AI affidabile**, al fine di promuovere la crescita inclusiva, lo sviluppo sostenibile e il benessere sociale delle persone, nel rispetto dello Stato di diritto, della trasparenza, dei diritti umani, dei valori democratici per garantire una società giusta e giusta. I sistemi di AI devono funzionare in modo solido, sicuro e protetto per tutto il loro ciclo di vita e **i potenziali rischi dovrebbero essere costantemente valutati e gestiti**. Le organizzazioni e le persone che sviluppano, implementano o gestiscono sistemi di AI dovrebbero essere ritenuti **responsabili** del loro corretto funzionamento in linea con i principi OCSE vigenti. **L'OCSE fornisce, inoltre apposite raccomandazioni ai governi**, sollecitati a facilitare gli investimenti pubblici e privati in ricerca e sviluppo per stimolare l'innovazione in un'AI affidabile, promuovere ecosistemi di intelligenza artificiale accessibili con infrastrutture

attività di sensibilizzazione *multi-stakeholder* e consultazione, anche con i governi, il mondo accademico, la società civile e il settore privato, non solo quelli nel G7 ma anche nelle economie esterne, comprese le economie in via di sviluppo ed emergenti. L'idea di fondo è quella di favorire un approccio aperto e abilitante ad un ambiente in cui i sistemi di AI siano sicuri, protetti, affidabili nonché progettati, sviluppati, distribuiti e utilizzati per massimizzare i benefici della tecnologia mitigandone i rischi, per il bene comune in tutto il mondo, anche nelle economie in via di sviluppo ed emergenti, con l'obiettivo di colmare il divario digitale e raggiungere l'inclusione digitale[26].

A **livello europeo** si è recuperata di recente la consapevolezza dell'esistenza di un *gap* con altre realtà geopolitiche. Dopo aver fissato le norme internazionali, dominato il progresso tecnologico e guidato la produzione e la diffusione di alta qualità, l'Europa è rimasta indietro, sviluppando e investendo assai meno nel mercato digitale di economie come gli Stati Uniti o la Cina, pur rimanendo relativamente competitiva nella produzione di ricerca tematica in materia di AI; è stato così riconosciuto il rischio che gli attori europei rimanessero emarginati nello sviluppo di standard globali e nei progressi della tecnologia e che gli stessi valori europei fossero messi in discussione.

L'Unione europea muove da una chiara consapevolezza delle prospettive di sviluppo, delle opportunità, delle sfide e dei rischi che l'AI impone agli attori pubblici e privati. Sotto quest'ultimo profilo è ormai acquisito che gli strumenti digitali stanno diventando sempre più un mezzo di manipolazione e abuso nelle mani di alcuni attori aziendali nonché di governi autocratici allo scopo di minare i sistemi politici democratici, portando così potenzialmente a uno scontro tra sistemi politici; lo spionaggio digitale, il sabotaggio, i conflitti su piccola scala e le campagne di

---

digitali in grado di condividere dati e conoscenze, garantire un ambiente politico favorevole alla distribuzione di sistemi di intelligenza artificiale affidabili, consentire alle persone di acquisire competenze in materia di AI, nonché cooperare a livello transfrontaliero e settoriale per progredire nella gestione responsabile di un'AI affidabile. In attuazione della Raccomandazione, è stato costituito l'**Osservatorio politico AI dell'OCSE** nel febbraio 2020, come hub innovativo con il compito di elaborare linee guida su metriche, politiche e pratiche di IA per aiutare a implementare i principi vigenti, facilitando il dialogo funzionale a condividere le migliori pratiche sulle politiche di IA, anche nella prospettiva di monitorare i progressi nella sua attuazione. Nel maggio 2020, durante la **riunione ministeriale** del **G7 sulla scienza e la tecnologia**, i Paesi del G7 hanno **formalizzato l'impegno di lanciare il partenariato globale sull'AI** per migliorare la cooperazione *multi-stakeholder* nel progresso dell'AI, aperta agli esperti e *stakeholder* di tutto il mondo, provenienti dal settore pubblico e privato, dalle comunità accademiche e scientifiche, nonché dalla società civile, per affrontare, secondo un approccio condiviso, le sfide globali legate allo sviluppo dell'intelligenza artificiale, con particolare riferimento alle principali priorità di intervento funzionali a favorire la ripresa economica nella fase post-epidemica, a seguito dell'emergenza sanitaria "Covid-19". Nel giugno 2019, il **G20 ha adottato i principi dell'AI incentrata sull'uomo** che si ispirano ai principi dell'AI dell'OCSE. In particolare, si è riconosciuta la **necessità di realizzare un modello sinergico di cooperazione** transnazionale nel settore dell'innovazione tecnologica, che persegua l'obiettivo di guidare lo sviluppo responsabile dell'AI nel rispetto dei diritti umani, inclusione, diversità e sostenibilità economica, fornendo supporto operativo teorico-pratico ad attività di ricerca all'avanguardia e a progetti di sperimentazione dei sistemi dell'IA.

[26] Tra i principi più significativi, da seguire secondo l'approccio basato sul rischio, sono indicati i seguenti: *i*) adozione delle misure adeguate per identificare, valutare e mitigare i rischi durante il ciclo di vita dell'AI; *ii*) monitoraggio delle vulnerabilità, degli incidenti, dei rischi emergenti e degli usi impropri dopo la distribuzione e adozione delle azioni appropriate per affrontarle; *iii*) segnalazione delle capacità, dei limiti e degli ambiti di applicazione dei sistemi avanzati di AI, per assicurarne la trasparenza e la responsabilità; *iv*) promozione della condivisione responsabile delle informazioni e della segnalazione degli incidenti tra organizzazioni che sviluppano sistemi avanzati di AI, anche con l'industria, con governi, con la società civile e con il mondo accademico; *v*) sviluppo, attuazione e divulgazione delle politiche di *governance* dell'AI e di gestione proporzionata del rischio, comprese le politiche sulla *privacy*, in particolare per le organizzazioni che sviluppano sistemi avanzati di IA; *vi*) investimento e implementazione di solidi controlli di sicurezza, inclusa la sicurezza fisica, la sicurezza informatica e le tutele contro le minacce interne durante tutto il ciclo di vita dell'AI; *vii*) sviluppo e implementazione di meccanismi affidabili di autenticazione e provenienza dei contenuti, per identificare i contenuti generati dall'AI; *viii*) attribuzione di priorità alla ricerca per mitigare i rischi sociali e per la sicurezza con definizione di priorità agli investimenti in misure di mitigazione efficaci; *ix*) attribuzione di priorità allo sviluppo di sistemi di AI avanzati per affrontare le più grandi sfide, in particolare, ma non solo, in relazione alla crisi climatica, alla salute globale e all'istruzione; *x*) promozione sello lo sviluppo e, ove appropriata, l'adozione di standard tecnici e migliori pratiche a livello internazionale; *xi*) implementazione delle misure adeguate di protezione dei dati personali e della proprietà intellettuale.

disinformazione sfidano le società democratiche. La natura dei modelli di *business* digitale consente un elevato grado di scalabilità ed effetti di rete; molti mercati digitali, infatti, sono caratterizzati da un alto grado di concentrazione del mercato, consentendo a un esiguo numero di piattaforme tecnologiche, attualmente aventi per lo più sede negli Stati Uniti, di guidare la commercializzazione di innovazioni tecnologiche pionieristiche, attrarre le migliori idee, i migliori talenti e le migliori imprese e conseguire una straordinaria redditività. Posizioni di mercato dominanti nell'economia dei dati possono estendersi all'economia emergente dell'AI; per contro, solo otto delle 200 principali aziende digitali odierne hanno sede nell'UE, mentre il completamento di un autentico mercato unico digitale si sta profilando come realtà della massima importanza e la competizione globale per la leadership tecnologica è diventata una priorità nell'UE, richiesta di agirà rapidamente e con coraggio, per evitare di dover seguire norme e standard fissati da altri, con effetti dannosi sulla stabilità politica, la sicurezza sociale, le libertà individuali e la competitività economica. Ancora, le tecnologie di AI rischiano di ridurre la capacità di intervento umano mentre l'AI dovrebbe rimanere una tecnologia antropocentrica affidabile, senza sostituire l'autonomia umana né importare la perdita della libertà individuale, ponendosi in termini inclusivi senza lasciare indietro nessuno.

In termini di benefici, l'AI è una delle tecnologie emergenti fondamentali della quarta rivoluzione industriale, utile per alimentare l'economia digitale, poiché consente l'introduzione di prodotti e servizi innovativi e ha il potere di aumentare la scelta dei consumatori o di rendere più efficienti i processi di produzione; secondo le previsioni, entro il 2030 l'AI contribuirà con oltre 11 miliardi di EUR all'economia globale. Le tecnologie di IA promettono di fornire un immenso valore economico alle economie che sviluppano in maniera redditizia, producono e adottano tali tecnologie, nonché alle economie e ai paesi in cui tale creazione di valore ha luogo; senza porsi quale tecnologia onnipotente, l'AI è uno strumento efficiente che può essere sfruttato a vantaggio della società ma in un quadro regolamentare capace di assicurare la protezione dei diritti e delle libertà fondamentali costruendo un AI quale effettivo uno strumento al servizio delle persone e della società, nel perseguimento del bene comune e dell'interesse generale, senza imporre ostacoli ingiustificati che impediscano agli attori europei di avere successo nell'era digitale, in particolare alle start-up e alle piccole e medie imprese (PMI). Il rapido progresso tecnologico introdotto dall'AI è sempre più inscindibile dalla maggior parte dei settori dell'attività umana e avrà ripercussioni anche sui mezzi di sussistenza di tutti coloro che non possiedono le competenze necessarie per adattarsi in modo sufficientemente veloce a tali nuove tecnologie; mentre il conseguimento dell'alfabetizzazione digitale mediante il miglioramento delle competenze e la riqualificazione può aiutare ad affrontare molte delle preoccupazioni socioeconomiche che ne derivano, tali impatti dovrebbero essere affrontati anche nel contesto dei sistemi di assistenza sociale, delle infrastrutture urbane e rurali e dei processi democratici.

L'approccio della UE all'AI è incentrato sull'eccellenza e sulla fiducia, con l'obiettivo di rafforzare la ricerca e la capacità industriale, garantendo la sicurezza e i diritti fondamentali. Vi è consapevolezza che il modo con il qual verrà fatto governo dell'AI definirà il mondo che viviamo in futuro. L'UE ha impostato una strategia volta alla costruzione di un'Europa resiliente per il decennio digitale che mira a renderla un polo di livello mondiale per l'AI, garantendo che la stessa resti incentrata sull'uomo e impegnata sull'affidabilità.

Nell'aprile 2021 la **Commissione** ha presentato il suo **pacchetto sull'IA**, composto dalla comunicazione sulla promozione di un approccio europeo all'IA[27], dalla revisione del piano coordinato sull'intelligenza artificiale (con gli Stati membri dell'UE)[28] e dalla proposta quadro di regolamentazione sull'AI e la pertinente valutazione d'impatto[29]. L'IA figura tra le principali priorità

[27] *Communication on Fostering a European approach to Artificial Intelligence,* reperibile su https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence.

[28] *Coordinated Plan on Artificial Intelligence,* reperibile su *https://digital-strategy.ec.europa.eu/en/policies/plan-ai*

[29] *Regulatory framework proposal on artificial intelligence,* reperibile su *https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.*

dell'attuale Commissione europea[30], alla ricerca di un approccio europeo coordinato alle implicazioni umane ed etiche e una riflessione volta a migliorare l'uso dei *big data* per favorire l'innovazione, garantendo la fiducia nelle tecnologie di AI senza pregiudizi per i diritti fondamentali dell'UE.

L'attenzione verso l'IA è sempre più sviluppata in ambito europeo; il **Parlamento europeo** che a far data dal 2016 ha emanato diverse risoluzioni in materia di *big data*, robotica e intelligenza artificiale, esaminando le implicazioni sollevate dall'IA e il modo in cui essa incide sul benessere, l'istruzione, la tecnologia, i diritti giuridici e fondamentali nonché l'industria in generale. Tali risoluzioni hanno sottolineato la necessità di adottare un approccio "antropocentrico" basato sul rispetto dei diritti fondamentali, riconosciuti dalla Carta dell'UE e dal quadro normativo dell'UE in materia di protezione dei dati[31]. Il Parlamento Europeo, nella raccomandazione alla Commissione

[30] Tra le comunicazioni della Commissione europea in tema di AI, si citano quelle di seguito indicate distinte per data e titolo: 25 aprile 2018 dal titolo "L'intelligenza artificiale per l'Europa" (COM(2018)0237); 7 dicembre 2018 dal titolo "Piano coordinato sull'intelligenza artificiale" (COM(2018)0795); 8 aprile 2019 dal titolo "Creare fiducia nell'intelligenza artificiale antropocentrica" (COM(2019)0168); 19 febbraio 2020 dal titolo "Una strategia europea per i dati" (COM(2020)0066); 19 febbraio 2020 dal titolo "Plasmare il futuro digitale dell'Europa" (COM(2020)0067); 10 marzo 2020 dal titolo "Una nuova strategia industriale per l'Europa" (COM(2020)0102) e del 5 maggio 2021 dal titolo "Aggiornamento della nuova strategia industriale 2020: costruire un mercato unico più forte per la ripresa dell'Europa" (COM(2021)0350); 30 settembre 2020 dal titolo "Piano d'azione per l'istruzione digitale 2021-2027 – Ripensare l'istruzione e la formazione per l'era digitale" (COM(2020)0624); 9 marzo 2021 dal titolo "Bussola per il digitale 2030: il modello europeo per il decennio digitale" (COM(2021)0118). A tali documenti si aggiungono: la **proposta di regolamento del Parlamento europeo e del Consiglio, del 21 aprile 2021**, che stabilisce regole armonizzate sull'intelligenza artificiale (**legge sull'intelligenza artificiale,** che identifica il cd. **AI act**) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206); la proposta di regolamento del Parlamento europeo e del Consiglio del 25 novembre 2020 relativo alla *governance* europea dei dati (Atto sulla governance dei dati) (COM(2020)0767); **il Libro bianco della Commissione del 19 febbraio 2020 dal titolo "Intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia" (COM(2020)0065)** e il Libro verde della Commissione del 27 gennaio 2021 dal titolo "Invecchiamento demografico. Promuovere la solidarietà e la responsabilità fra le generazioni" (COM(2021)0050).

[31] Vastissimo il quadro di regolamenti, risoluzioni e direttive del Parlamento europeo su temi che afferiscono all'AI. Tra essi, si ricorda: i) il regolamento (UE) 2021/694 del Parlamento europeo e del Consiglio, del 29 aprile 2021, che istituisce il programma Europa digitale e abroga la decisione (UE) 2015/2240(5); il regolamento (UE) 2021/695 del Parlamento europeo e del Consiglio, del 28 aprile 2021, che istituisce il programma quadro di ricerca e innovazione Orizzonte Europa e ne stabilisce le norme di partecipazione e diffusione, e che abroga i regolamenti (UE) n. 1290/2013 e (UE) n. 1291/2013; il regolamento (UE) 2018/1807 del Parlamento europeo e del Consiglio, del 14 novembre 2018, relativo a un quadro applicabile alla libera circolazione dei dati non personali nell'Unione europea; il regolamento (UE) 2021/697 del Parlamento europeo e del Consiglio, del 29 aprile 2021, che istituisce il Fondo europeo per la difesa e abroga il regolamento (UE) 2018/1092(8); la direttiva (UE) 2019/770 del Parlamento europeo e del Consiglio, del 20 maggio 2019, relativa a determinati aspetti dei contratti di fornitura di contenuto digitale e di servizi digitali; il regolamento (UE) 2021/1173 del Consiglio del 13 luglio 2021 relativo all'istituzione dell'impresa comune per il calcolo ad alte prestazioni europeo e che abroga il regolamento (UE) 2018/1488(10). Numerose infine le risoluzioni sull'argomento in analisi come quelle di seguito elencate; risoluzione del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica; risoluzione del 1° giugno 2017 sulla digitalizzazione dell'industria europea; risoluzione del 6 ottobre 2021 sul quadro strategico dell'UE per la sicurezza stradale 2021-2030 – Raccomandazioni sulle prossime tappe verso l'obiettivo "zero vittime"; risoluzione del 12 settembre 2018 sui sistemi d'arma autonomi; risoluzione del 12 febbraio 2019 su una politica industriale europea globale in materia di robotica e intelligenza artificiale (IA); risoluzione del 12 febbraio 2020 dal titolo "Processi decisionali automatizzati: garantire la tutela dei consumatori e la libera circolazione di beni e servizi"; risoluzione del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale; risoluzione del 20 ottobre 2020 sui diritti di proprietà intellettuale per lo sviluppo di tecnologie di intelligenza artificiale; risoluzione del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate; risoluzione del 20 gennaio 2021 sull'intelligenza artificiale: questioni relative all'interpretazione e applicazione del diritto internazionale nella misura in cui l'UE è interessata relativamente agli impieghi civili e militari e all'autorità dello Stato al di fuori dell'ambito della giustizia penale; risoluzione del 20 maggio 2021 dal titolo "Plasmare il futuro digitale dell'Europa: eliminare gli ostacoli al funzionamento del mercato unico digitale e migliorare l'uso dell'IA per i consumatori europei"; risoluzione del 25 marzo 2021 su una strategia europea per i dati; risoluzione del 19 maggio 2021 sull'intelligenza artificiale nell'istruzione, nella cultura e nel settore audiovisivo; **risoluzione del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale**; **la risoluzione del Parlamento europeo del 3 maggio 2022 sull'intelligenza artificiale in un'era digitale.** Si

Europea sulla finanza digitale (2020/2034) ha affermato che «La finanza digitale svolge pertanto un ruolo fondamentale nello sviluppo delle attività finanziarie e nell'incremento della loro diffusione, dato che la sua applicazione comporta benefici significativi, tra cui un generale vantaggio in termini di efficienza, riduzioni dei costi, migliore tutele per i consumatori e una migliore gestione dei dati e trasparenza»., il Parlamento Europeo ha dato mandato alla Commissione di "sviluppare un approccio proporzionato, trasversale e olistico nei confronti delle sue attività in materia di *FinTech*, traendo insegnamenti da quanto è stato fatto in altre giurisdizioni e adeguandosi alla diversità degli operatori e dei modelli imprenditoriali impiegati" e di rivedere la legislazione in materia di servizi finanziari al fine di renderla "sufficientemente favorevole all'innovazione finanziaria". L'obiettivo è piuttosto chiaro: predisporre un *framework* normativo che sia *innovation-friendly*, neutrale e adatto alle nuove realtà importate dal Fintech.

Va poi ricordata la **Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi** adottata dalla CEPEJ nel corso della sua 31ª Riunione plenaria di Strasburgo, del 3-4 dicembre 2018. La Carta è destinata agli attori pubblici e privati incaricati di creare e lanciare strumenti e servizi di IA relativi al trattamento di decisioni e dati giudiziari (apprendimento automatico o qualsiasi altro metodo derivante dalla scienza dei dati), oltre che ai responsabili di decisioni pubbliche competenti in materia di quadro legislativo o regolamentare e ai soggetti addetti allo sviluppo, alla verifica o all'utilizzo di tali strumenti e servizi.

Prendendo atto della crescente importanza dell'IA nelle moderne società nonché dei benefici previsti quando sarà utilizzata pienamente al servizio dell'efficienza e della qualità della giustizia, la CEPEJ ha adottato formalmente i cinque principi fondamentali (la cd. "Carta etica europea") sull'utilizzo dell'IA nei sistemi giudiziari e negli ambiti connessi: *i*) il principio del rispetto dei diritti fondamentali nell'elaborazione e nell'attuazione di strumenti e servizi di IA; *ii*) il principio della non-discriminazione, dovendosi prevenire specificamente lo sviluppo o l'intensificazione di discriminazioni tra persone o gruppi di persone; *iii*) il principio di qualità e sicurezza, che richiede, in ordine al trattamento di decisioni e dati giudiziari, l'utilizzo di fonti certificate e dati intangibili, con modelli elaborati interdisciplinarmente, in un ambiente tecnologico sicuro; *iv*) il principio di trasparenza, imparzialità e equità, per il quale occorre rendere le metodologie di trattamento dei dati accessibili e comprensibili, autorizzando verifiche esterne; *v*) il principio "del controllo da parte dell'utilizzatore", dovendosi precludere un approccio prescrittivo e assicurare che gli utilizzatori siano attori informati e abbiano il controllo delle loro scelte.

L'utilizzo degli strumenti e dei servizi di IA nei sistemi giudiziari è finalizzato a migliorare l'efficienza e la qualità della giustizia e merita di essere incoraggiato, chiarendo che va svolto in modo responsabile, nel dovuto rispetto dei diritti fondamentali della persona, enunciati nella Convenzione europea sui diritti dell'uomo e nella Convenzione per la protezione dei dati di carattere personale, e in conformità agli altri principi fondamentali che devono orientare la definizione delle politiche pubbliche in materia di giustizia anche in questo campo.

Quanto agli **strumenti utilizzati dalle autorità investigative prima del processo penale**, oggetto più immediato di interesse di questa riflessione, la Carta riconosce che gli strumenti denominati di "polizia predittiva" (prima del processo giudiziario o del rinvio a giudizio) sono già in rapida crescita e cominciano a essere noti al grande pubblico; si pensi alla lista di interdizione al volo (*no fly list*), applicazione dell'analisi dei megadati che raccoglie e analizza dati riguardanti potenziali terroristi al fine di prevenire la commissione di atti, o agli algoritmi utilizzati per scoprire le frodi o il riciclaggio di denaro; si utilizzano correntemente un gran numero di strumenti informatici per prevenire la commissione di reati (mediante l'individuazione dei possibili luoghi in cui ciò potrebbe

richiama anche la proposta di decisione del Parlamento europeo e del Consiglio del 15 settembre 2021 che istituisce il programma strategico per il 2030 "Percorso per il decennio digitale" (COM(2021)0574).

avvenire o i loro autori) o per perseguirli in maniera più efficace[32] (§ 120 della Carta etica).

La prima categoria comprende gli strumenti di "polizia predittiva" utilizzati per prevenire alcune tipologie di reato realizzati con regolarità, quali il furto con effrazione, la violenza di strada, il furto di veicoli o di oggetti situati al loro interno; essi mirano a potenziare la capacità di determinare con precisione dove e quando tali reati potrebbero essere commessi e a riprodurre tali informazioni su una carta geografica sotto forma di "punti caldi" sorvegliati in tempo reale da pattuglie della polizia, secondo un processo denominato "mappatura dei rischi di reato" (*predictive criminal mapping*). La maggior parte dei *software* utilizzati in tale ambito si fonda su elementi di localizzazione storica dei reati forniti dai rapporti di polizia, ma sono in fase di sperimentazione nuove tecnologie che combinano vari dati provenienti da fonti differenti[33].Tali strumenti, provvisti di tassi di efficacia convincenti, avrebbero anche effetti dissuasivi in ordine alla commissione di reati nelle zone circostanti i punti segnalati[34]. Nondimeno le capacità predittive di questi strumenti mostrano anche limiti. Alcuni di questi ultimi si rivelano in relazione ai reati di natura meno regolare o che colpiscono luoghi diversi, come appunto gli atti di terrorismo, altri ambiti di interesse di questo elaborato. Un altro punto debole è l'effetto dei "circoli viziosi" e delle "profezie che si auto-adempiono": i quartieri considerati a rischio attirano maggiormente l'attenzione della polizia, la quale scopre conseguentemente un maggior numero di reati, con il risultato di un'eccessiva sorveglianza da parte della polizia delle comunità residenti in tali luoghi[35]. Infine, negli stessi servizi di polizia non mancano interrogativi su una possibile "tirannia dell'algoritmo" che potrebbe minimizzare o addirittura sostituire progressivamente il giudizio umano anche se, allo stato, la tecnologia è presentata come al servizio degli esseri umani.

Resta riconoscibile la tendenza ad intensificare l'**applicazione dell'analisi dei megadati nel contrasto e nel perseguimento dei reati**. Strumenti quali *Connect*, utilizzato dalla polizia del Regno Unito per analizzare miliardi di dati generati dalle transazioni finanziarie al fine di trovare correlazioni o schemi operativi, o la banca dati ICSE (*International Child sexual Explotation Database*, ovvero la banca dati internazionale sullo sfruttamento sessuale dei minori, gestita dall'Interpol, che aiuta a identificare le vittime e gli autori di reati analizzando, per esempio, l'arredamento o altri oggetti presenti nelle immagini degli abusi, o ancora l'analisi nei video di voci sullo sfondo), si sono dimostrati particolarmente efficaci nella lotta contro i reati[36]. Ciononostante, la dottrina si interroga sulla logica manageriale della risposta alla criminalità fornita da tali strumenti predittivi, in cui l'analisi approfondita delle ragioni alla base del reato diviene meno importante rispetto all'intervento sul posto e immediato. Ciò accade in un momento in cui le risorse finanziarie

[32] A. ZAVRSNIK, *Big Data, crime and social control*, 017, pp. 194 e ss. elenca in modo dettagliato una serie di strumenti utilizzati dai servizi di polizia in Europa e negli Stati Uniti d'America.

[33] Per esempio, nel quadro del progetto "*E-Security - ICT for knowledge-based and predictive urban security*" (http://www.esecurity.trento.it/), svolto nella città italiana di Trento, tra il novembre 2012 e il maggio 2015, una banca dati che raccoglie informazioni sui reati denunciati alla polizia, i risultati di inchieste condotte dall'amministrazione comunale sulla vittimizzazione e sulla sicurezza reale e percepita dei cittadini, informazioni sul disordine urbano materiale e sociale provenienti dalla polizia, nonché altre variabili relative alla "*SmartCity*" (per esempio informazioni sul contesto socio-demografico, sul contesto urbano, sull'illuminazione notturna, sulla presenza di videocamere di sorveglianza e del trasporto pubblico). Tale banca dati è stata creata per fornire strumenti migliori per l'opera di prevenzione dei reati e di miglioramento della sicurezza urbana. I responsabili del progetto hanno testimoniato l'affidabilità delle tecniche utilizzate, che consentono asseritamente di prevedere la commissione di reati con un tasso di attendibilità pari a circa il 60-65 % e sarebbero inoltre in grado di migliorare la lotta al crimine in presenza di risorse limitate. Inoltre, esperimenti svolti nel Regno Unito nel quadro di un progetto pilota finalizzato a individuare anticipatamente i luoghi di possibili furti, furti con effrazione, e aggressioni, indicano che le proiezioni del software utilizzato, denominato PREDPOL, erano accurate nel 78% dei casi rispetto al 51 % delle tecniche tradizionali.

[34] L'indicazione della concentrazione geografica dei reati consentirebbe alle forze di polizia di valutare in modo migliore i fattori ambientali che rendono più probabile la commissione di reati nella zona esaminata (illuminazione, presenza di esercizi commerciali, ecc.) e di pianificare, di concerto con altri partner, risposte adeguate.

[35] "*Predicting crime, LAPD style*", The Guardian, 25 giugno 2014.

[36] Grazie a *Connect*, per esempio, ricerche che richiedevano precedentemente mesi di indagini, possono essere svolte attualmente in pochi minuti in relazione a una notevole quantità di dati e con un elevato livello di complessità.

disponibili sono in calo e la polizia, pur essendo dotata di personale, attrezzature e risorse limitate, deve fornire lo stesso livello di protezione pubblica.

Il **14 giugno 2023** il Parlamento europeo ha approvato in sede plenaria la bozza del c.d. *Artificial Intelligence Act* (o "AI Act"), primo testo di legge per disciplinare e l'uso dell'IA a livello globale; tale approvazione apre alla negoziazione con il Consiglio dell'Unione europea sul testo definitivo del regolamento che rappresenterà uno *standard* a livello mondiale per lo sviluppo e la *governance* dell'IA. La bozza di regolamento ha ampliato l'elenco dei divieti sugli usi intrusivi e discriminatori dell'AI, tra cui: - l'uso per il c.d. *social scoring* (vala a dire la classificazione delle persone in base al loro comportamento sociale o alle loro caratteristiche personali); - l'uso di sistemi di identificazione biometrica remota "in tempo reale" e "a posteriori" in spazi accessibili al pubblico; - i sistemi di polizia predittiva (basati su profilazione, ubicazione o comportamenti criminali passati); - l'uso di *software* quale sistema di riconoscimento delle emozioni utilizzato dalle forze dell'ordine nella gestione delle frontiere; - l'estrazione non mirata di dati biometrici da internet o da filmati di telecamere a circuito chiuso per creare database di riconoscimento facciale. I sistemi di AI generativa (come, ChatGPT) dovrebbero, tra gli altri, rispettare i requisiti di trasparenza (dichiarando che il contenuto è stato generato dall'AI), aiutando anche a distinguere le immagini c.d. *deep-fake* da quelle reali e fornire salvaguardie per evitare la generazione di contenuti illegali. Dovrebbero inoltre essere rese pubbliche le sintesi dettagliate dei dati protetti dal diritto d'autore utilizzati per l'addestramento. Sono infine previste esenzioni per le attività di ricerca e le componenti dell'AI fornite con licenze *open-source*. All'elenco dei sistemi di AI "a rischio elevato"[37] si aggiungono quelli che possono provocare danni alla salute, alla sicurezza o ai diritti fondamentali; alla stregua di questo criterio aggiuntivo, sono considerati ad alto rischio anche i sistemi di raccomandazione delle piattaforme *online* di grandi dimensioni, come definiti dal *Digital Services Act*. L'AI Act detta le regole per lo sviluppo, l'immissione sul mercato e l'utilizzo di sistemi di IA nell'Unione europea. Trattandosi di un regolamento, le sue norme si applicheranno in modo uniforme e diretto in tutti gli Stati membri, senza necessità di leggi di recepimento nazionali. Il legislatore europeo ha scelto un approccio orizzontale, dunque non settoriale, costruito su una nuova definizione di «sistema di IA» basato sul rischio. Vengono così vietate le pratiche di intelligenza artificiale a rischio inaccettabile, come l'IA che attribuisce un punteggio sociale. Per i sistemi ad alto rischio – ad esempio quelli impiegati per la valutazione degli studenti – sono imposti diversi requisiti e obblighi, mentre per quelli a rischio limitato, come i sistemi di *deep fake*, si prevedono specifiche prescrizioni per garantirne la

[37] I sistemi di IA sono considerati a **rischio inaccettabile**, e pertanto vietati, quando costituiscono una minaccia per le persone. Questi comprendono la manipolazione comportamentale cognitiva di persone o gruppi vulnerabili specifici (ad esempio giocattoli attivati vocalmente che incoraggiano comportamenti pericolosi nei bambini), la classificazione sociale (classificazione delle persone in base al comportamento, al livello socio-economico, alle caratteristiche personali), i sistemi di identificazione biometrica in tempo reale e a distanza, come il riconoscimento facciale Alcune eccezioni potrebbero tuttavia essere ammesse: per esempio, i sistemi di identificazione biometrica a distanza "post", in cui l'identificazione avviene dopo un significativo ritardo, saranno consentiti per perseguire reati gravi e solo previa autorizzazione del tribunale. I sistemi di IA che influiscono negativamente sulla sicurezza o sui diritti fondamentali saranno considerati ad **alto rischio** e saranno suddivisi in due categorie: 1) I sistemi di intelligenza artificiale utilizzati in prodotti soggetti alla direttiva dell'UE sulla sicurezza generale dei prodotti. Questi includono giocattoli, aviazione, automobili, dispositivi medici e ascensori. 2) I sistemi di intelligenza artificiale che rientrano in otto aree specifiche dovranno essere registrati in un database dell'UE: - identificazione e categorizzazione biometrica di persone naturali; gestione e funzionamento di infrastrutture critiche; istruzione e formazione professionale; occupazione, gestione dei lavoratori e accesso all'autoimpiego; accesso e fruizione di servizi privati essenziali e servizi pubblici e vantaggi; forze dell'ordine; gestione delle migrazioni, asilo e controllo delle frontiere assistenza nell'interpretazione e applicazione legale della legge. Tutti i sistemi di intelligenza artificiale ad alto rischio saranno valutati prima di essere messi sul mercato e durante tutto il loro ciclo di vita. L'**IA generativa**, come ChatGPT, dovrà rispettare requisiti di trasparenza: - rivelare che il contenuto è stato generato da un'intelligenza artificiale; - progettare il modello in modo da impedire la generazione di contenuti illegali; - pubblicare riepiloghi dei dati con diritti d'autore utilizzati per l'addestramento. I sistemi di intelligenza artificiale a **rischio limitato** dovrebbero rispettare requisiti minimi di trasparenza che consentano agli utenti di prendere decisioni informate. Dopo aver interagito con le applicazioni, l'utente può decidere se desidera continuare a utilizzarle. Gli utenti dovrebbero essere informati quando interagiscono con l'IA. Questo include i sistemi di IA che generano o manipolano contenuti di immagini, audio o video (ad esempio *deepfake*).

trasparenza. L'AI Act consentirà anche di adottare codici di condotta per l'applicazione volontaria dei requisiti fissati per i sistemi di IA ad alto rischio. Sono inoltre stabilite misure per favorire l'innovazione, come ad esempio la creazione di spazi di sperimentazione normativa o alcune agevolazioni per startup e PMI.

Il negoziato tra le istituzioni europee è in corso e sono diversi i punti del testo ancora da definire. Si prevede, in ogni caso, che il regolamento possa essere approvato tra la fine del 2023 e primi mesi del 2024. Ciò anche in vista delle elezioni europee del prossimo giugno. Tuttavia, le norme dell'*AI Act* non si applicheranno immediatamente. È infatti previsto un periodo cuscinetto di 24 mesi (nelle versioni di Commissione e Parlamento) o di 36 mesi (nella posizione del Consiglio). Pertanto, a parte alcune specifiche disposizioni ad applicazione anticipata e salve successive modifiche, il regolamento non dovrebbero applicarsi prima del 2026.

- 4. Le nozioni di base.

Insopprimibili le esigenze sottese alla **questione definitoria**. L'esame di un fenomeno che non muova da definizione condivise rischia di condurre l'analisi su sentieri autoreferenziali, nei quali ciascuno matura convinzioni soggettive e non costruisce una condivisione di conoscenze, base per un'evoluzione comune. Nella materia che impegna notevoli sono gli **ostacoli** al raggiungimento di una condivisione estesa sul contenuto e la struttura di alcuni dei termini che evocano fenomeni di ricorrente impiego; primo fra tutti, come di vedrà, il termine "intelligenza artificiale". L'originalità e la complessità dei nuovi schemi tecnologici e informatici, in continua evoluzione e in costante affinamento, la ricorrente fallace impostazione esposta al pericolo di scambiare l'obiettivo con la realtà acquisita, l'auspicio con la proprietà già presente. E nondimeno è dalla costruzione di definizioni comuni che passa una partecipazione più consapevole e responsabile dei diversi attori del nuovo sistema.

Come ricorda la Commissione europea nel libro bianco sull'intelligenza artificiale "Un approccio europeo all'eccellenza e alla fiducia"[38] «qualunque nuovo strumento giuridico dovrà comprendere una definizione di IA abbastanza **flessibile** da accogliere il progresso tecnico, ma anche sufficientemente precisa da garantire la necessaria **certezza** del diritto». In questa prospettiva, che può apparire limitata, ma che è assai realistica in merito ai rapporti costantemente tesi della tecnologia con il diritto, a seguire si ripercorrono alcuni dei **tentativi definitori** di maggior rilievo ai fini che interessano.

- 4 .1. L'intelligenza artificiale: tra miti, realtà e limiti.

- 4.1.1. I tentativi definitori.

Nella sua **comunicazione** dal titolo «L'intelligenza artificiale per l'Europa» la Commissione europea ha fornito una prima definizione dell'IA; in particolare, tale espressione «indica **sistemi** che mostrano un **comportamento** intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi. I sistemi basati sull'IA possono consistere solo in *software* che agiscono nel mondo virtuale (ad esempio assistenti vocali, *software* per l'analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale), oppure incorporare l'IA in dispositivi *hardware* (per esempio in *robot* avanzati, auto a guida autonoma, droni

[38] Cfr. COM (2020) 65 final, Bruxelles, 19.2.2020.

o applicazioni dell'Internet delle cose[39])»[40]. Questa definizione è stata ulteriormente perfezionata dal **gruppo di esperti** ad alto livello che hanno definito l'IA nei seguenti termini: «I sistemi di intelligenza artificiale (IA) sono sistemi *software* (ed eventualmente *hardware*) progettati dall'uomo che, dato un obiettivo complesso, agiscono nella dimensione fisica o digitale percependo il proprio ambiente attraverso l'acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, ragionando sulle conoscenze, o elaborando le informazioni derivate da questi dati e decidendo le migliori azioni da intraprendere per raggiungere l'obiettivo dato. I sistemi di IA possono usare regole simboliche o apprendere un modello numerico, e possono anche adattare il loro comportamento analizzando come l'ambiente è influenzato dalle loro azioni precedenti»[41].

Ai fini del **libro bianco sull'IA** e di eventuali future discussioni su iniziative strategiche, la Commissione ha ritenuto opportuno chiarire gli elementi principali da cui è composta l'IA, vale a dire i "**dati**" e gli "**algoritmi**"[42]. L'IA può essere integrata nell'*hardware*. Nel caso delle tecniche di apprendimento automatico, che costituiscono un sottoinsieme dell'IA, gli algoritmi vengono addestrati per dedurre determinati modelli partendo da un *set* di dati al fine di stabilire le azioni necessarie al conseguimento di un determinato obiettivo[43]. Gli algoritmi possono continuare a imparare mentre vengono utilizzati. Sebbene i prodotti basati sull'IA possano agire in modo autonomo percependo il proprio ambiente senza seguire un insieme di istruzioni predeterminate, il loro comportamento è in larga misura definito e limitato dai loro sviluppatori. Gli esseri umani determinano e programmano gli obiettivi che un sistema di IA dovrebbe raggiungere nel modo più efficace.

Si tratta di definizione assai generica che identifica l'IA nell'«insieme di tecnologie che combina dati, algoritmi **e potenza di calcolo**», riconnettendo ai progressi compiuti nell'ambito del calcolo e alla crescente disponibilità di dati i fattori determinanti per la crescita dell'IA[44]. Questa definizione rimarca che l'IA si basa sulla raccolta, l'analisi e l'accumulo ricorrente di ingenti quantità di **dati**, compresi quelli personali, provenienti da fonti diverse, oggetto di trattamenti automatizzati mediante algoritmi informatici e altre tecniche avanzate che utilizzano sia dati memorizzati sia in *streaming*, per individuare correlazioni, tendenze e modelli (analisi dei *Big Data*). La definizione mostra di considerare l'IA come una delle più importanti applicazioni dell'economia dei dati,

[39] *Internet of Things* (IoT) è la rete globale di tutti i dispositivi e delle macchine abilitate connessi a Internet che possono raccogliere, inviare, condividere e agire sui dati, utilizzando sensori, processori e hardware di comunicazione incorporati, senza interazione umana. Tale rete di oggetti fisici, ossia le "things", con sensori, software e altre tecnologie integrate allo scopo di connettere e scambiare dati con altri dispositivi e sistemi su Internet, vanno dai normali oggetti domestici ai sofisticati strumenti industriali. L'IoT genera un'enorme quantità di dati in tempo reale che possono essere analizzati e utilizzati per creare le azioni desiderate o i risultati aziendali.

[40] Cfr. COM (2018) 237 final, pag. 1.

[41] Gruppo indipendente di esperti sull'intelligenza artificiale, istituito dalla Commissione europea nel giugno 2018, 8 aprile 2019, pag. 8.

[42] La parola "algoritmo" è stata coniata dal matematico Muhammad Ibn Musa, nel IX secolo, per designare qualunque schema o procedimento sistematico di calcolo; è l'espressione matematica del concetto di procedura generale, di metodo sistematico valido per la soluzione di una classe di problemi identificando una sequenza di istruzioni elementari e ripetibili che consente di risolvere un problema. Un algoritmo, quindi, richiede sempre una serie di passaggi per risolvere un quesito che viene scomposto in calcoli elementari.

[43] Nel caso della guida autonoma, ad esempio, l'algoritmo utilizza in tempo reale dati forniti dall'automobile (velocità, consumo del motore, dati relativi agli ammortizzatori, ecc.) e dai sensori che rilevano tutti i dati dell'ambiente in cui si trova l'automobile (strada, segnali, altri veicoli, pedoni, ecc.) per stabilire quale direzione dovrebbe prendere l'automobile e con quale accelerazione e velocità al fine di raggiungere una determinata destinazione. L'algoritmo si basa sui dati rilevati e si adegua alla situazione della strada ed alle condizioni esterne, compreso il comportamento degli altri conducenti, al fine di ottenere la guida più sicura e più comoda. Nel caso della guida autonoma, ad esempio, l'algoritmo utilizza in tempo reale dati forniti dall'automobile (velocità, consumo del motore, dati relativi agli ammortizzatori, ecc.) e dai sensori che rilevano tutti i dati dell'ambiente in cui si trova l'automobile (strada, segnali, altri veicoli, pedoni, ecc.) per stabilire quale direzione dovrebbe prendere l'automobile e con quale accelerazione e velocità al fine di raggiungere una determinata destinazione. L'algoritmo si basa sui dati rilevati e si adegua alla situazione della strada ed alle condizioni esterne, compreso il comportamento degli altri conducenti, al fine di ottenere la guida più sicura e più comoda.

[44] Cfr. COM (2020) 65 final, Bruxelles, 19.2.2020.

enfatizzando l'attenzione sulla **risorsa** gestita (il "petrolio dei dati") e sullo **strumento** impiegato per raffinarla (conducendo a dati nuovi). Tali dati provengono dagli individui ma anche dalle stesse applicazioni di IA che, a loro volta, si alimentano di dati – personali e non - provenienti dall'industria, dalle imprese e dal settore pubblico, trattati per scopi diversi, con decisioni automatizzate con effetti sugli individui, a iniziare dai principi fondamentali della protezione dei dati e della vita privata. Si tratta di una **definizione di contesto** che indica in termini volutamente generici alcuni contenuti di partenza e gli strumenti dell'AI (dati e calcoli), ma che non chiarisce in dettaglio il metodo di correlazione tra dati già conosciuti e dati "scoperti" (la cd. nuova conoscenza).

La proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'Unione [45]fissa regole armonizzate per lo sviluppo, l'immissione sul mercato e l'utilizzo di sistemi di IA nell'Unione seguendo un approccio proporzionato basato sul rischio. Essa propone un'unica definizione di IA adeguata alle esigenze future. I portatori di interessi hanno richiesto per lo più una definizione restrittiva, chiara e precisa del concetto di intelligenza artificiale. I portatori di interessi hanno altresì sottolineato che, oltre al chiarimento del termine "intelligenza artificiale", è importante definire anche "rischio", "alto rischio", "basso rischio", "identificazione biometrica remota" e "danno". La definizione di sistema di IA nel quadro giuridico mira ad essere il più possibile neutrale dal punto di vista tecnologico e adeguata alle esigenze future, tenendo conto dei rapidi sviluppi tecnologici e di mercato relativi all'IA. Al fine di fornire la necessaria certezza del diritto, il titolo I è integrato dall'allegato I, contenente un elenco dettagliato di approcci e tecniche per lo sviluppo dell'IA che deve essere adattato dalla Commissione in linea con i nuovi sviluppi tecnologici. Tra i considerando risulta che la nozione di sistema di IA dovrebbe essere definita in maniera chiara per garantire la certezza del diritto, prevedendo allo stesso tempo la flessibilità necessaria per agevolare i futuri sviluppi tecnologici. La definizione dovrebbe essere basata sulle principali caratteristiche funzionali del software, in particolare sulla **capacità, per una determinata serie di obiettivi definiti dall'uomo, di generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano l'ambiente con cui il sistema interagisce**, tanto in una dimensione fisica quanto in una dimensione digitale. I sistemi di IA possono essere progettati per funzionare con **livelli di autonomia variabili** e per essere utilizzati come e**lementi indipendenti (stand-alone) o come componenti di un prodotto**, a prescindere dal fatto che il sistema sia fisicamente incorporato nel prodotto (integrato) o assista la funzionalità del prodotto senza esservi incorporato (non integrato). La definizione di sistema di IA dovrebbe essere completata da un elenco di tecniche e approcci specifici utilizzati per il suo sviluppo, che dovrebbe essere tenuto aggiornato alla luce degli sviluppi di mercato e tecnologici mediante l'adozione da parte della Commissione di atti delegati volti a modificare tale elenco. Secondo l'art. 3, comma 1, della proposta di regolamento il sistema di intelligenza artificiale" (sistema di IA) è un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono.

La definizione può essere posta in relazione – tra le molte – con quella fornita dal Consiglio di Stato italiano (Sez. III, sent. 25 novembre 2021 n. 7891), che lega il concetto a quello di *Machine Learning* (ML). In questa più restrittiva concezione, in contrapposizione all'algoritmo "tradizionale", l'IA sarebbe connotata dalla capacità di elaborare costantemente nuovi criteri di inferenza tra dati e assume[re] decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico.

L'IA è indubbiamente concetto molto lato e dai confini non ancora condivisi in seno alla Comunità internazionale, sovente frainteso con strumenti di alto impatto informatico.

Lo stesso **Parlamento europeo**, rispondendo all'interrogativo di cosa sia l'IA, fuori da contesti

[45] Cfr. COM (2021) 206 final, Bruxelles, 21.4.2021

normativi ufficiali, la definisce (cfr. informativa su sito *internet* al 13 giugno 2023) quale «l'abilità di una macchina di mostrare capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività» e di permettere «ai sistemi di capire il proprio ambiente, mettersi in relazione con quello che percepisce e risolvere problemi, e agire verso un obiettivo specifico. Il computer riceve i dati (già preparati o raccolti tramite sensori, come una videocamera), li processa e risponde. I sistemi di IA sono capaci di adattare il proprio comportamento analizzando gli effetti delle azioni precedenti e lavorando in autonomia». L'illustrazione appare coerente con una logica comunicativa di massa orientata alla semplificazione ma non riveste, come si diceva, una portata ufficiale e propriamente definitoria. In quest'ultima prospettiva, ad esempio, sempre il Parlamento europeo[46] ha identificato il sistema di intelligenza artificiale (IA) nel «un sistema basato su software o integrato in dispositivi hardware che mostra un comportamento che simula l'intelligenza, tra l'altro raccogliendo e trattando dati, analizzando e interpretando il proprio ambiente e intraprendendo azioni, con un certo grado di autonomia, per raggiungere obiettivi specifici». In realtà che l'IA simuli o sia in grado di simulare quella umana è affermazione assai complessa come si vedrà in seguito, mentre che essa persegua "obiettivi specifici" e con un grado di autonomia limitato corrisponde più ad auspicio che a condizione reale.

Per altra definizione, che aspira a maggiore specificità, l'"Intelligenza artificiale" identifica in generale diverse teorie, metodologie e tecniche che consentono di progettare soluzioni informatiche in grado di riprodurre a vario titolo l'intelligenza umana[47] profilandosi come «la branca della scienza che studia la realizzazione di macchine che abbiano l'abilità di mostrare capacità umane, quali il ragionamento, l'apprendimento, la pianificazione e la creatività, imitando in qualche modo il comportamento umano»[48].

Si tratta di un ramo dell'informatica confrontatosi, negli anni, con alcune intrinseche limitazioni dell'evoluzione tecnologica, in termini di capacità di calcolo, di elaborazione di grandi quantità di dati, di maturità degli algoritmi. La rapida evoluzione del comparto tecnologico ha reso via via disponibili le necessarie capacità elaborative in grado di sostenere gli algoritmi di AI, che quindi – grazie anche all'ampia disponibilità di dati – hanno iniziato a rilasciare le loro potenzialità nel tessuto produttivo (si pensi all'applicazione di tecniche di *image detection* nelle filiere produttive ai fini di controllo di qualità) ma anche nella sfera individuale (si consideri la diffusione di interfacce conversazionali, quali *chatbot*, *voicebot* o assistenti virtuali, anche su dispostivi mobili)[49]. Queste tecniche si differenziano significativamente dall'approccio tradizionale della programmazione, nel quale un programmatore definisce un algoritmo in grado, attraverso la codifica di una serie di operazioni, di trasformare deterministicamente dei dati di *input* nei dati di *output* desiderati. Nell'AI, in luogo della formalizzazione di un algoritmo deterministico si costruisce un **modello** (secondo approcci induttivi o deduttivi, v. *infra*) in grado di catturare le informazioni necessarie per derivare "**conoscenza**" in modo automatico a partire dai dati disponibili. In questo ambito, le competenze di sviluppo delle applicazioni vengono integrate con quelle tipiche di *data scientist*, figura professionale orientata all'analisi dei dati e con significative conoscenze del dominio di *business*.

Per l'interesse più immediato di questo approfondimento, nella relazione sulle opportunità e le sfide delle nuove tecnologie per il contrasto del riciclaggio e del finanziamento del terrorismo del luglio 2021[50], il GAFI identifica l'IA nella **scienza** che imita le capacità di pensiero umano per eseguire compiti quali il riconoscimento di modelli, la formulazione di previsioni e di

[46] Risoluzione del 2021 su *Intelligenza artificiale: questioni relative all'interpretazione e applicazione del diritto internazionale (al di fuori del processo penale).*
[47] Cfr. Banca d'Italia
[48] J. MARCUCCI, *op. cit., passim.*
[49] Per una ricognizione dell'uso dell'AI nel settore bancario cfr. CIPA-ABI (2021), *Rilevazione sull'IT nel settore bancario italiano.*
[50] FATF (2021), *Opportunities and Challenges of New Technologies for AML/CFT*, FATF, Paris, France, *https://www.fatf-gafi.org/publications/ fatfrecommendations/documents/opportunities-challenges-new- technologies-aml-cft.html*

raccomandazioni o l'assunzione di decisioni utilizzando tecniche computazionali avanzate per ottenere intuizioni (insight) da dati conseguiti da fonti di diversa tipologia e qualità (strutturate e non strutturate) in modo da risolvere problemi ed eseguire compiti con diversi livelli di autonomia, attraverso sistemi che combinano intenzionalità, intelligenza e adattabilità. Nel glossario predisposto per la medesima relazione, un sistema di IA è quello basato su macchina che può, per un predeterminato insieme di obiettivi definiti dall'uomo, fare previsioni, proporre raccomandazioni o assumere decisioni che influenzano ambienti reali o virtuali e operare con diversi livelli di autonomia[51]. Al momento, l'apprendimento automatico è la forma di intelligenza artificiale più familiare e sviluppata. L'obiettivo dell'IA è consentire ai computer di automatizzare alcuni aspetti dell'analisi, riservando all'intervento umano compiti più sottili e di ottenere conoscenze che gli esseri umani potrebbero non raggiungere. Oltre al fatto che l'IA funziona con schemi tecnologici diversi e numerose applicazioni, la problematicità della definizione scaturisce, ancor prima, dal fatto che non sono condivise le nozioni di "pensiero", di "intelligenza" o di "completa autonomia".

- **4.1.2. Il metodo induttivo e l'approccio deduttivo.**

L'AI si può sinteticamente ripartire in due differenti impostazioni, da cui derivano diverse tecniche algoritmiche: l'approccio **induttivo** e l'approccio **deduttivo**.

Nell'approccio induttivo la macchina sintetizza la propria conoscenza sulla base dell'osservazione empirica dei dati, "imparando" da questi tramite un processo di generalizzazione, ossia traendo dai dati particolari una regola generale; questo approccio è conosciuto come apprendimento automatico, o ***Machine learning*** (ML).

Nell'approccio deduttivo, invece, partendo da una rappresentazione formale della conoscenza operata tramite linguaggi di rappresentazione e ragionamento della conoscenza (*knowledge representation and reasoning*, KRR), la macchina produce nuova conoscenza dai dati in *input* seguendo un processo di inferenza che dalla regola generale conduce al fatto particolare; è il cd. **ragionamento automatico** (o *Automated reasoning*, AR**)**.

Volendo esemplificare l'applicazione dei due approcci all'AI, gli stessi si riflettono un diverso orientamento alla soluzione dei problemi. Ipotizzando un problema di decisione da affidare alla macchina (la necessità di decidere in merito ad una certa situazione non nota a priori, come, ad esempio, svoltare ad un'uscita autostradale o proseguire il viaggio, concedere o negare un credito). Adottando un metodo induttivo, la macchina, precedentemente addestrata su un ampio numero di coppie problema-decisione (ad esempio un insieme di scelte di viaggio o di decisioni su specifiche richieste di credito), sarà in grado di prendere la decisione astraendo dai casi concreti analizzati, grazie a modelli statistici in grado di approssimare le dinamiche della decisione sulla base delle caratteristiche della nuova situazione. Con un metodo deduttivo, invece, la macchina, precedentemente istruita sulle dinamiche del dominio di interesse (la regola di come ci si comporta in autostrada all'approssimarsi della destinazione o secondo quali criteri si concede o si nega un credito), applicherà al caso specifico le nozioni apprese, fornendo la risposta.

Il mutamento di paradigma portato dall'intelligenza artificiale è, in ogni caso, travolgente. Nell'informatica tradizionale, i problemi vengono risolti mediante tecniche di natura deterministica: si svolge un'analisi del problema dato (e/o dei requisiti), si implementa un algoritmo che produce il risultato finale (*output)* previsto a seguito di esperienze ripetute nel tempo. Con l'IA, invece, si adotta un approccio differente, di natura induttiva (come con il *machine learning*) o deduttiva (con i sistemi di ragionamento automatico): in entrambi i casi abbiamo sistemi che apprendono informazioni in ingresso (*input)* e sono in grado di determinare e inferire conoscenza in modo automatico e

[51] OECD (2020), *AI Principles, https://www.oecd.ai/ai-principles*.

semiautomatico[52]. Le soluzioni informatiche e gli algoritmi sono maturi da un punto di vista tecnico: ve ne sono alcuni più di frontiera, altri più consolidati, ma in questo ambito si può ormai sfruttare un'ampia esperienza riveniente anche dal mondo accademico; nondimeno l'incertezza non sembra ovviabile dai sistemi di IA e questo è un aspetto che segna una netta linea di demarcazione con l'informatica tradizionale: le soluzioni dotate di elementi di AI, infatti, presentano rischi peculiari, come la possibilità di produrre risultati non utili o non corretti, non avere una quantità o una qualità) sufficiente di dati in *input* da cui produrre nuova e qualificata conoscenza[53].

- **4.1.3. I limiti reali dei sistemi di IA.**

Merita qui dar conto di alcuni **limiti degli attuali sistemi di AI;** limiti non è ancora ben chiaro quanto possano ritenersi superabili e che danno anche ragione delle ricorrenti osservazioni per cui un'intelligenza artificiale è strutturalmente deficitaria rispetto ai caratteri identitari dell'intelligenza almeno secondo gli archetipi umani.

Sul piano della logica, infatti, i tipi di ragionamento possibili agli esseri umani sono tre, non potendosi ridurre alla deduzione e all'induzione. Viene in rilievo, infatti, anche l'**abduzione,** che è la forma di ragionamento caratteristica degli uomini ed ancora irriproducibile dalle macchine, utile una conoscenza frutto di creatività, intuizione, fiducia; tutte espressioni che nella prospettiva delle prime potrebbe corrispondere più ad errori o difetti di funzionamento che a criteri di risoluzione di problemi. Sulla strada di una IA "integrata", oltre a nuove connessioni tra *deduzione* (applicazione di regole ai dati) e *induzione* (costruzione di regole dai dati), si dovranno fare passi importanti verso procedimenti di carattere semiotico come quelli dell'abduzione. La differenza tra deduzione, induzione e abduzione si illustra con l'esempio del sacco di fagioli di Peirce. Si immagini un tavolo su cui si trova un sacco di fagioli. Se sul sacco è posta l'etichetta "fagioli neri", si può *dedurre* che prendendo una manciata di fagioli dal sacco questi siano neri. Se il sacco non reca alcuna etichetta ma prendendo una manciata si osserva che tutti fagioli sono neri, si può allora *indurre* (eventualmente dopo un certo numero di manciate) che tutti i fagioli nel sacco siano neri. Se il sacco reca l'etichetta "fagioli neri" e una manciata di fagioli neri fuori dal sacchetto, posata sul tavolo, si può pensare che i fagioli provengano da quel sacco è questa un'*abduzione*, ossia **una ragionevole ipotesi su una situazione osservata, un ragionamento sulla migliore spiegazione dei fatti.** Gran parte del *senso comune* che applichiamo nella vita quotidiana, *in primis* quello che ci consente di comprendere il linguaggio, ricade sotto questa specie. Ora, l'**IA ha sinora maggiormente sviluppato metodi deduttivi e induttivi, meglio trattabili in termini matematici. L'abduzion**e, infatti, **avviene naturalmente nella mente umana senza precise garanzie di tipo logico o empirico**. Al pari di un'arte o comunque di momenti in cui l'essere umano applica la sua libera intenzionalità. I ragionamenti sulle migliori spiegazioni possono avvalersi di molte euristiche ma non hanno una forma universale, dipendono cioè da **modelli specifici ed estesi**, su grandi depositi di conoscenze, come quelle che consentono le diagnosi mediche. L'IA integrata consisterà in una sinergia tra metodi di apprendimento necessari per costruire modelli di grande estensione e metodi di ragionamento ipotetico basato su logiche non classiche (ad esempio probabilistiche) per applicare tali modelli ai casi.

Resta il fatto che l'IA, nelle attuali configurazioni, si basa su varie forme di induzione e deduzione, ossia sui metodi di ragionamenti che ci fanno imparare da ripetizione, ricorsività, analisi, statistica, esperienza e derivazioni. Lo stesso concetto di rete neuronale ha ampliato lo spettro di tali processi ma non ne ha alterato la natura. Il progetto di IA di creare una macchina abduttiva è ancora lontano dall'essere attuato. Eppure, è proprio **l'abduzione uno dei metodi tipici dell'indagine in ambito investigativi o penale, assistendo il passaggio dal conseguente all'antecedente**, quale

[52] F. FEDERICO, *AI e prospettive di evoluzione dell'architettura informatica*, in F. FEDERICO F, MARCUCCI J, BEVILACQUA M, MARCHETTI DJ, *Rapporto 2/2021 – L'impiego dell'intelligenza artificiale nell'attività di Banca d'Italia,* in BioLaw, 24 dicembre 2021.
[53] F. FEDERICO, *AI e prospettive*

processo che serve agli esseri umani per formulare un'ipotesi di fronte a un fenomeno mai incontrato in precedenza cambiando totalmente il genere o il *framework* dell'ipotetica regola che serve a spiegare il fenomeno. Charles S. Peirce, usava come esempio quello degli "Assassini della Rue Morgue" di E. A. Poe: un omicidio troppo violento e senza spiegazione valida degli indizi viene risolto uscendo dal paradigma "essere umano come possibile assassino", leggendo i caratteri estetici degli indizi come segno della colpevolezza di un orango-tango. La stessa **creatività** nella lettura dei segni va ascritta a molte delle scoperte umane, da quella dell'orbita ellittica dei pianeti di Keplero a quella della penicillina di Fleming, nonché a quella che ci fa produrre una scultura o una poesia. Nell'esperienza quotidiana tale lettura dei segni si rivela alla decisione istantanea di fidarsi di uno sconosciuto o si applica alla prima infanzia contraddistinta da grandi incontri con esperienze del tutto ignote. L'IA potrà andare molto lontano arrivando a straordinari risultati deduttivi e induttivi, formulare anche ipotesi non totalmente creative che sono sofisticazioni di questi processi e che spesso affollano l'esistenza umana, ma difficilmente riuscirà a essere abduttiva, cioè a cambiare del tutto il piano di spiegazione, da quello fisico della faccia dello sconosciuto a quello morale della fiducia, per esempio. La creatività artificiale, dunque, decisiva e importantissima, mantiene una struttura di genere induttivo o deduttivo: è destinata ad ampliare enormemente la conoscenza ma difficilmente, si crede, giungerà e varcherà mai la soglia della creatività umana.

Occorre inoltre riconoscere che i sistemi di IA implementano le condizioni della conoscenza dell'uomo, ma non consentono la comprensione profonda di essa senza l'apporto critico dello stesso. Va ricordata, in proposito, la differenza tra **IA debole e IA forte**. L'IA debole identifica la capacità di una macchina di implementare una parte di «intelligenza» per eseguire un compito preciso; è questa l'IA con cui abbiamo a che fare quotidianamente (es. di pensi all'assistente vocale, al sistema di riconoscimento immagini, etc.). L'**IA Forte (o IA generale)** identifica, invece, la capacità di una macchina di comprendere o apprendere ogni tipo di compito «intellettuale» che un essere umano è in grado di comprendere o apprendere; in altre parole, una macchina che ha coscienza di sé. Questo approdo non è attuale ed è dubitabile lo sarà mai.

La comprensione profonda del linguaggio è forse la sfida più grande che l'IA si troverà ad affrontare nei prossimi anni. L'IA intellettiva ha conseguito spettacolari successi in alcuni tipi di analisi linguistica, come quella sintattica. Nessuno potrebbe proporre oggi un analizzatore sintattico (*parser*) che non sia basato su reti neurali addestrate su vasti *corpora* testuali. Sul versante della semantica[54], l'IA *data-driven* (ri)propone l'analisi distribuzionale, basata sulla classica idea che le parole che occorrono nello stesso contesto abbiano qualcosa a che fare l'una con l'altra sul piano del significato (se si dice "vado al cinema" e "vado a teatro", allora "cinema" e "teatro" hanno qualche relazione semantica). Con questo approccio si può in effetti giungere a valutare con buona accuratezza la similarità di due frasi, rimanendo tuttavia totalmente all'oscuro del loro significato e delle loro implicazioni. Il **muro del significato** non è ancora abbattuto nonostante le aspettative riposte nei nuovi assistenti virtuali. La **comprensione profonda del linguaggio è forse il terreno sul quale l'integrazione delle due anime dell'IA è lontana da prospettive di successo.** Che i modelli neurali del linguaggio stentino ad elevarsi al livello della semantica, specie sotto il profilo della composizionalità, e che sia necessario far in qualche modo ricorso alla conoscenza di senso comune, è ormai posizione comune nel **dibattito scientifico** sull'IA. Per vedere qualcosa che somiglia a una *comprensione profonda* del linguaggio all'opera nei nostri elettrodomestici bisognerà però aspettare ancora qualche tempo.

Altro tema ricorrente nelle definizioni, poi, è quello dello ***human in the loop***, cioè del necessario controllo **umano**. Tuttavia, se questa può essere una prescrizione normativa, certamente essa non corrisponde alla realtà della IA e della sua potenzialità e appare di difficile applicabilità, visto il consenso non universale su tale condizionalità. Tuttavia, anche tale ultimo approccio non è

[54] La semantica è il ramo della linguistica che si occupa dei fenomeni del significato del linguaggio e non dal punto di vista fonetico e morfologico.

privo di problemi definitori, in quanto la costante progressione della capacità di ML, unita alla disponibilità di una quantità enorme e sempre crescente di dati e alla capacità computazionale dei nuovi strumenti, determinano problemi del tutto nuovi anche per ciò che concerne l'individuazione del rapporto causale e del principio di responsabilità. La novità delle questioni poste dalla IA, intesa propriamente nei termini sopra indicati, è resa ancora più evidente dalla diffusione del *Quantum Computing*, cioè dell'applicazione delle metodiche computazionali quantistiche, che si basano non su inferenze causali deterministiche, ma su logiche meramente probabilistiche. Essa è considerata una delle principali minacce alla stabilità internazionale, secondo plurime fonti. Inoltre, il ritmo dello sviluppo della IA vede ormai una progressione geometrica. Lo sviluppo dell'IA generalista, ad esempio, è individuato tra le principali sfide esistenziali raccolte nel *GlobalTrends* per il 2040 del *National Intelligence Council* statunitense.

- **4.2. Il *machine learning* e i diversi approcci all'apprendimento automatico**.

Il *Machine Learning* è un tipo (o sottoinsieme o branca) di IA che "addestra" i sistemi informatici ad apprendere dai dati, a identificare da essi modelli e prendere decisioni con un intervento umano minimo. L'apprendimento automatico implica la progettazione di una sequenza di azioni per risolvere un problema automaticamente attraverso l'esperienza e l'evoluzione di algoritmi di riconoscimento dei modelli con intervento umano limitato o nullo, ML è dunque è un metodo di analisi dei dati che automatizza la costruzione di modelli analitici, per lo più, per identificare degli andamenti (o *patterns*) in grosse moli di dati e fare previsioni[55]. A differenza di un sistema che esegue un'attività seguendo regole esplicite e predefinite (sistema cd. *rule-based*), un modello di machine learning **apprende costantemente dall'esperienza**. Mentre un sistema basato su regole predeterminate esegue un'attività **ogni volta allo stesso modo** (nel bene o nel male), le **prestazioni** (*performance*) di un sistema di ML possono essere **migliorate attraverso l'apprendimento**, grazie all'applicazione di **metodi statistici**, esponendo l'algoritmo a una maggiore quantità di dati differenti

Per contro, questo significa che **i risultati di un modello di machine learning non sono mai completamente certi** e vanno considerati con cautela, tanto che sovente i risultati sono corredati da una percentuale di accuratezza stimata, rimettendo all'utilizzatore, in base alla criticità e al rischio correlato a una sbagliata interpretazione, stabilire la percentuale di confidenza minima da utilizzare.

Il mutamento di paradigma è assolutamente travolgente: secondo l'impostazione informatica tradizionale un problema può essere affrontato con un approccio di sviluppo del software, che consiste nell'**identificare e scrivere una funzione specifica** in base a cui un certo input produce sempre un certo output. In un problema risolto dalla funzione y=f(x), è il programmatore a descrivere nel dettaglio come funziona "f(x)". Con il machine learning, invece, si usano **algoritmi matematici e statistici generici** che, esposti a una determinata serie di dati in una fase iniziale definita "di **addestramento**" (*training*) e passando per una seconda fase di valutazione dei risultati con ottimizzazione dei parametri, ricavano autonomamente la funzione – non sempre conosciuta e non sempre conoscibile dallo sviluppatore – in grado di individuare in una **differente serie di dati** (dati di esecuzione), il **valore più probabile di y**, indicando eventualmente un grado di confidenza nella stima. In pratica, ricava da solo la funzione f(x). Il sistema composto da **algoritmo addestrato**, **dati** e **parametri operativi** viene chiamato **modello.** Diversi sono gli algoritmi per risolvere differenti problemi e le fasi di training e valutazione dei modelli. Resta che i modelli di machine learning sono molto efficaci nell'individuare **caratteristiche comuni o tendenze in enormi serie di dati**, prendendo in considerazione un **numero di variabili che nessun essere umano può essere in grado di valutare, o addirittura di notare.**

[55] J. MARCUCCI, *Big Data, Machine Learning e Artificial Intelligence nell'analisi economica e statistica, in* in F. FEDERICO F, MARCUCCI J, BEVILACQUA M, MARCHETTI DJ, *Rapporto 2/2021 – L'impiego dell'intelligenza artificiale nell'attività di Banca d'Italia,* in *BioLaw*, 24 dicembre 2021.

Nella *machine learning (ML)* esistono diversi approcci all'apprendimento automatico, che determinano le caratteristiche degli algoritmi e i requisiti richiesti (ad es. in termini di numerosità dei dati necessari per l'addestramento). Gli approcci più noti sono quello supervisionato, non supervisionato, semi-supervisionato, con rinforzo.

Nell'approccio ***supervisionato***, l'algoritmo apprende il modello delle relazioni tra gli input e gli output mediante un insieme di dati preventivamente etichettati da un essere umano (cosiddetto *labeled dataset*); i dati di input dell'algoritmo sono costituiti da risultati noti e gli algoritmi sono addestrati tramite l'esempio. La coppia input/output (dati etichettati) fornisce *riscontri* per l'algoritmo, che utilizza il set di dati di addestramento per regolare il modello e ridurre al minimo l'errore. Ad esempio, un set di addestramento può contenere immagini di diversi tipi di animali con associate etichette, consentendo all'algoritmo di confrontare l'etichetta prevista con quella corretta. In altri termini, i dati con cui gli algoritmi vengono istruiti sono già contrassegnati con risposte valide, o in ogni caso è possibile valutare i risultati per confermare o correggere il modello. Per esempio, per istruire un algoritmo su come riconoscere immagini di gatti, gli si forniscono una serie di immagini di gatti a cui si applica l'etichetta "è un gatto", e immagini di oggetti o animali diversi con l'etichetta "non è un gatto". L'apprendimento supervisionato utilizza un set di dati di convalida per misurare i progressi dell'algoritmo nell'apprendimento del modello e un set di dati di test per valutare le prestazioni del modello su dati mai visti prima per determinare se il modello ha appreso i dati di *training* in modo efficace ed è in grado di generalizzare nuovi dati.

Nell'approccio ***semi-supervisionato*** il *dataset* è solo parzialmente etichettato (ad es. perché il costo di *labeling* di un *dataset* è solitamente alto); infine, in quello "***con rinforzo"*** l'algoritmo compie azioni in modo da massimizzare progressivamente una funzione di profitto, che assegna un valore positivo o negativo ad ogni azione, secondo una tecnica che si potrebbe definire "*trial and error*". Nell'approccio ***non supervisionato***, l'algoritmo apprende autonomamente il modello dal *dataset*, senza necessità che questo sia anticipatamente processato per l'attribuzione delle *label (etichette)*; l'apprendimento non supervisionato è dunque un processo di apprendimento automatico che consente agli algoritmi di analizzare e raggruppare set di dati non etichettati (con la risposta corretta) per scoprire modelli nascosti e non conosciuti in precedenza, raggruppamenti di dati o anomalie o caratteristiche (pattern) senza intervento umano. L'algoritmo analizza i dati disponibili e determina correlazioni e relazioni senza una chiave di risposta traendo inferenze e raggruppando cose simili sulla base di osservazione e intuizione non vincolate. Man mano che la quantità di dati a cui è esposto l'algoritmo cresce, la modellazione diventa più accurata e raffinata.

L'approccio supervisionato viene spesso usato nei problemi di classificazione, dove si beneficia di *dataset* già provvisti di una corretta assegnazione in classi che l'algoritmo può apprendere ed applicare sui nuovi dati. Queste tecniche riescono a produrre modelli efficaci con volumi di dati tendenzialmente inferiori rispetto all'approccio non supervisionato, usato quando si vuole apprendere la struttura inerente dei dati senza partire da una conoscenza formalizzata degli stessi. Tra i problemi governabili con sistemi di ML supervisionati vi sono, oltre a quelli di classificazione (previsione di risposte non numeriche, come la probabilità di un mancato pagamento del mutuo) anche quelli di **regressione** (previsione di risposte numeriche, come il numero di prodotti che verranno venduti il mese prossimo in un punto vendita di Roma). Nel ML non supervisionato, invece, si possono distinguere problemi di **clustering o raggruppamento** (ricerca di gruppi di oggetti simili, come scarpe da corsa, scarpe da passeggio e scarpe eleganti per cui l'algoritmo è in grado di partizionare i dati in insiemi contenenti informazioni simili tra loro), **associazione** (ricerca di sequenze comuni di oggetti, come caffè e latte) e di **riduzione della dimensionalità** (o *dimensionality reduction*, cioè la proiezione, selezione ed estrazione delle caratteristiche chiave di un modello, per esempio allo scopo di eliminare da una tabella le colonne con i dati non necessari o fuorvianti).

Esistono, inoltre, due famiglie di **algoritmi**, applicabili trasversalmente negli approcci sopracitati, significativi per la loro diffusione o le loro intrinseche caratteristiche.

La prima va sotto il nome di "apprendimento di insieme" (o "***ensemble learning***")[56]**.** Le tecniche di *ensemble learning* prevedono, per lo svolgimento dell'esercizio di previsione, l'utilizzo di insiemi di modelli in quanto la previsione finale è generalmente ottenuta come la previsione media o maggioritaria di essi. Un caso molto diffuso è quello in cui i singoli modelli sono alberi decisionali (cfr. esempio in figura 2) che operano tramite una partizione progressiva dello spazio delle soluzioni. Le tecniche possono essere suddivise in due sotto-categorie basate su *bagging* (apprendimento simultaneo di modelli indipendenti tra loro, ciascuno caratterizzato dallo sfruttamento di porzioni dell'informazione complessiva) e su *boosting* (sequenze di modelli che vanno via via a raffinare il processo di apprendimento). Due popolari esempi per gli approcci citati, nel contesto dell'apprendimento supervisionato, sono, rispettivamente, il *random forest* e il *gradient boosting*.

Ulteriore famiglia di algoritmi di ML è il ***Deep learning*** nel quale i processi elaborativi, ispirandosi al comportamento dei neuroni del cervello umano, si basano su reti che interconnettono nodi organizzati in livelli successivi (cd. reti neurali); ad ogni livello della rete corrisponde una fase dell'apprendimento di concetti via via sempre più complessi. Il DL è una forma avanzata di ML in cui reti neurali artificiali con numerosi strati (profondi) apprendono da grandi quantità di dati in modo altamente autonomo. Gli algoritmi DL eseguono un compito ripetutamente, ogni volta modificandolo leggermente per migliorare il risultato, consentendo alle macchine di risolvere problemi complessi senza l'intervento umano. In tali casi sono utilizzate strutture e algoritmi molto più complicati, come le reti neurali artificiali, caratterizzati da numerosi strati interconnessi tra loro, per risolvere problemi molto complessi, come ad esempio un sistema di guida autonoma.

- **4.3. Big data, advanced analytics, analisi testuali e AI.**

Il dominio dell'IA è sovente accomunato all'uso dei c.d. **big data**, che richiamano una prospettiva diversa anche se con rilevanti punti di contatto.

La definizione di big data è molto ampia e può dipendere dal contesto. Gran parte di questi dati è in una forma di tipo non strutturato, ossia una forma che ne impedisce la rappresentazione attraverso una tabella in un classico database relazionale. Spesso tali dati non strutturati sono in forma testuale e necessitano di tecniche elaborazione del linguaggio naturale *(natural language processing*, NLP), che sono algoritmi di IA in grado di analizzare, rappresentare e quindi comprendere il linguaggio naturale.

Con il termine ***Big data*** solitamente si individua l'utilizzo di algoritmi e altre soluzioni per l'analisi di grandi quantità di dati memorizzati in archivi eterogenei e non collegati tra di loro. A differenza dei sistemi tradizionali di gestione dati, i *Big data* comprendono anche dati semi-strutturati e non strutturati (es. commenti su social networks, tracce audio-visive). Più esattamente i Big data si riferiscono sia a dati contraddistinti da alcune caratteristiche descritte nel seguito, sia all'insieme di algoritmi, tecnologie e soluzioni informatiche in grado di offrire servizi di raccolta, gestione e analisi degli stessi.

Il paradigma associato ai *big data* è quello delle cosiddette "**5 V**", che individuano tre

[56] L'**apprendimento d'insieme** (*ensemble learning*) indica una serie di metodi che usano molteplici modelli o algoritmi per ottenere una migliore prestazione predittiva rispetto a quella ottenuta dagli stessi modelli applicati singolarmente. A differenza dell'insieme della meccanica statistica, che si ritiene infinito, tale insieme di modelli alternativi è concreto e finito. L'apprendimento d'insieme si divide in tre tecniche fondamentali. *Bagging* è la tecnica che mira a creare un insieme di classificatori aventi la stessa importanza; all'atto della classificazione, ciascun modello voterà circa l'esito della predizione e l'*output* complessivo sarà la classe che avrà ricevuto il maggior numero di voti; nel *Boosting* ciascun classificatore influisce sulla votazione finale con un certo peso, calcolato in base all'errore di accuratezza che ciascun modello commetterà in fase di learning; *Stacking* introduce un ulteriore classificatore (detto meta-classificatore) che utilizza le predizioni di altri sotto-modelli per effettuare un ulteriore learning (nel *bagging* l'output era il risultato di una votazione).

caratteristiche distintive e due prassi attese:

- **volume**: i dati disponibili per attività di analisi spesso si attestano nell'ordine dei *terabyte* o superiori e dunque è richiesta una grossa mole di dati;

- **varietà**: riferita ai formati, ai più tradizionali dati di natura strutturata, in questo paradigma si affiancano anche quelli semi-strutturati (come file XML) o non strutturati (come documenti testuali o immagini);

- **velocità**: i dati sono prodotti a ritmi estremamente elevati[57], pertanto occorre dotarsi di tecnologie in grado di processarli con adeguata velocità ovvero tecniche adeguate alla loro analisi in tempo reale;

- **veridicità**: poiché i dati possono essere affetti da inattendibilità, dovuta alla natura dei processi di generazione e di raccolta delle osservazioni, occorre fare in modo che gli stessi rappresentino quanto più possibile fedelmente la realtà sottostante;

- **Valore**: occorre essere in grado di trasformare il dato in informazione utile al *business*.

Pertanto, si definiscono *big data* gli insiemi di osservazioni che presentino almeno una tra le caratteristiche di alto volume (nel numero di osservazioni o nel numero di attributi), alta varietà (di contenuto o formato) e velocità di produzione o raccolta, tali da implicare il ricorso a strumenti e tecniche non tradizionali. Perciò, possono afferire alla categoria di *big data* sia grandi volumi di transazioni e pagamenti, connotati da alta granularità di informazioni, che dati in formato testuale, come ad esempio le causali di spesa o bonifico, come pure i dati di fonte *social network* e quelli connessi alla navigazione su *internet*.

I *big data* sono, dunque, il risultato dell'accumulazione di tutta una serie di dati legati alle varie attività degli individui, registrati attraverso sensori o *app* istallate sugli *smartphone* o le attività effettuate sul *web*, come ad esempio la scrittura di e-mail o di opinioni su un servizio o un prodotto acquistato, l'attività sui *social networks* come *Twitter* e *Facebook*. Tali dati vengono spesso prodotti da società private e utilizzati principalmente per il loro *business*, ma possono essere utilizzati anche dagli istituti di statistica, dalle agenzie governative e dalle banche centrali per ottenere degli indicatori economici più ad alta frequenza, al fine di capire l'andamento dell'economia in momenti caratterizzati da particolari shock come la recente pandemia da COVID-19. In tali momenti, infatti, l'attività degli istituti di statistica può essere fortemente limitata, determinando l'assenza totale o parziale di informazioni tipicamente ottenute con le indagini e la conseguente mancata produzione di indicatori economici e statistici tradizionali.

Il legame tra il dominio dei *big data* e le tecniche di AI è immediato: la disponibilità di insiemi di dati di dimensioni molto rilevanti unita a quella delle tecnologie per elaborarli in modo efficiente consentono agli algoritmi di apprendimento automatico di "imparare" meglio. Tali algoritmi, però, possono essere addestrati, anche in assenza dei *big data* se si dispone di *dataset* di dimensioni e caratteristiche adeguate allo scopo; onde la presenza di big data non è un prerequisito per l'applicazione di tecniche di AI. Per contro, numerose tecniche (che ricadono nel dominio a cui ci si riferisce comunemente col termine *analytics*) applicabili ai *big data* che non prevedono l'impiego di algoritmi di AI. I due domini, pertanto, hanno un rapporto di complementarietà, beneficiando reciprocamente dei progressi compiuti.

I grossi volumi e la natura non strutturata dei *big data* rappresentano anche un limite per coloro

[57] Come ricorda J. MARCUCCI, *op. cit., passim.*, su internet in un minuto, nell'anno 2021, si sono registrati circa 200.000 individui che scrivono tweets su Twitter, circa 200 milioni di emails inviate, circa 70 milioni di messaggi inviati tra WhatsApp e Messenger, 10.000 connessioni su Linkedin o 21 milioni di messaggi di di testo inviati per telefono. Dati ottenuti da Bond High Plus sul sito web https://bit.ly/3suyIa3.

che vogliano utilizzarli, *richiedendo* piattaforme di calcolo distribuito specifiche (i c.d. *data lakes),* che permettono la convivenza e lo sfruttamento di dati tradizionali e non tradizionali. Inoltre, sono richiesti *skills* specifici e necessarie competenze come quelle del *data scientist*, ossia di un esperto di molti domini (informatica, economia, statistica, ingegneria); è inoltre, fondamentale una strategia di cooperazione orizzontale, tra le diverse aree organizzative, sia tra le risorse umane, facendo collaborare gli «esperti di dominio» con competenze diverse, ma con una profonda conoscenza del *business* di interesse per l'azienda (come economisti, statistici, avvocati, ecc.) con gli ingegneri, gli informatici, i fisici e i matematici per modellare e programmare questi algoritmi di intelligenza artificiale e machine learning), si potrà lavorare proficuamente sui big data[58].

Tra le applicazioni dell'IA vi sono anche gli **approcci analitici avanzati** (*advanced analytics*) ovvero un insieme di tecniche, autonome o semi- autonome, che permettono di sviluppare modelli prescrittivi e predittivi. Un sistema di analisi avanzata è basato su tecniche complesse di *machine learning* e di analisi autonoma o semi-autonoma dei dati ed è volto a prevedere tendenze future e risultati di business e migliorare l'analisi tradizionale. Tramite l'analisi in tempo reale, la modellazione predittiva e metodi statistici l'*advanced analytics* permette di prevedere scenari e tendenze future. Tali sistemi e strumenti vanno oltre lo studio del dato storico, cercando le possibili correlazioni in ottica predittiva (fornendo predizioni future su eventi noti o sconosciuti, basandosi su dati storici e attuali, tramite tecniche di modellazione predittiva, *machine learning* e *data mining*) e addirittura prescrittiva (partendo dall'analisi descrittiva e predittiva, specifici software suggeriscono quali siano le migliori azioni da intraprendere per raggiungere determinati obiettivi, mostrando anche gli effetti derivanti da ogni decisione). Molteplici strumenti analitici avanzati per analisi dei megadati e supporto decisionale avanzato sono a disposizione delle forze dell'ordine e possono aiutarli nelle loro indagini.

Completando l'esposizione delle ultime nozioni preliminari per l'esame del tema oggetto di più diretto interesse, con l'espressione "**analisi testuale**" si richiama l'utilizzo di algoritmi ed altre soluzioni per l'estrazione, l'analisi e la classificazione di informazioni da documenti testuali non strutturati, includendo soluzioni di *Text mining* e *Natural Language Processing* (NLP). L'"**analisi di rete**" identifica l'insieme di metriche (es. densità, direzionalità delle relazioni) utilizzate descrivere e analizzare le principali caratteristiche di una rete, ovvero di un insieme di entità (es. persone fisiche, imprese, transazioni) connesse tra di loro.

La ***Blockchain*** o altre soluzioni di *Distributed Ledger Technology* identificano la tecnologia che permette la creazione e gestione di un registro distribuito in cui tutti i nodi della rete contribuiscono al mantenimento della sua integrità e le transazioni tra i nodi della rete vengono autenticate tramite l'utilizzo di chiavi crittografiche.

Con l'espressione ***Cloud computing*** si richiamano le soluzioni tecnologiche messe a disposizione da un fornitore di servizi che permettono l'accesso da remoto, tramite la rete internet, a risorse *hardware* e/o *software* per l'elaborazione dei dati. Questa definizione include tutte le modalità di accesso all'infrastruttura *cloud* (*cloud* pubblico, *cloud* privato e *cloud* ibrido).

Le "**tecnologie biometriche"** infine, identificano soluzioni tecnologiche che permettono l'identificazione, la verifica delle generalità e l'assegnazione delle credenziali d'autenticazione al cliente tramite l'utilizzo di una o più caratteristiche biologiche del soggetto (es. impronta digitale, riconoscimento facciale, riconoscimento vocale).

- 4.4 Explainable AI e black box.

L'applicazione di tecniche di AI importa aspetti non consueti nello sviluppo di soluzioni informatiche tradizionali: tra questi, la spiegabilità dei risultati prodotti dall'algoritmo, che si pone

[58] J. MARCUCCI, *op. cit., passim.*

alla base della capacità, da parte dell'analista, di illustrare agli *stakeholder* del processo le motivazioni a supporto delle decisioni prese (o suggerite) dall'algoritmo.

Con il termine *explainable AI* (XAI) si indica l'insieme di strumenti applicati dall'analista volti ad integrare il risultato primario dell'algoritmo di ML con un insieme di interpretazioni e spiegazioni (*explanation*) circa il funzionamento del modello. Da un punto di vista concettuale, l'XAI può riferirsi a due approcci principali: - l'**interpretabilità** (*interpretability*), tracciando quantitativamente i meccanismi che governano il comportamento del modello; - la **spiegabilità** (*explainability*), formulando valutazioni qualitative (giustificazioni) sui risultati ottenuti, per spiegare le logiche di funzionamento del modello. Il grado di spiegabilità intrinseca dei diversi algoritmi che compongono il panorama delle tecniche di *machine learning* tende di norma a diminuire con l'aumento della capacità previsiva, presentando un rapporto che in generale è di proporzionalità inversa: ad esempio le reti neurali offrono *performance* molto elevate e possono essere rese in qualche misura spiegabili solo adottando tecniche XAI; sul fronte opposto un modello intrinsecamente spiegabile, come l'albero decisionale, presenta di norma minore capacità previsiva. Su un piano realizzativo, le tecniche di XAI si distinguono tra quelle che possono essere applicate solo a specifiche classi di modelli (la cosiddetta *model-specific explainability*), come la *feature importance* degli alberi decisionali, e quelle applicabili a qualsiasi modello a prescindere dalla sua forma funzionale (cosiddetta *model-agnostic explainability*), come SHAP[59] e LIME[60] . L'implementazione di tecniche di XAI può essere radicata in fasi diverse del ciclo di vita dei modelli e più precisamente: prima dello sviluppo del modello, con riguardo alla comprensione dei dati usati per addestrarlo (*pre-modeling explainability*); durante lo sviluppo del modello, per favorirne l'interpretabilità (*explainable modeling*); a valle dello sviluppo, per spiegare a posteriori la logica decisionale del modello (*post-modeling, o post hoc explainability*).

Nel contesto delle nuove tecnologie, la spiegabilità significa che i processi, le soluzioni o i sistemi basati sulla tecnologia possono essere spiegati, compresi e contabilizzati. Tale condizione ricorre quando è offerta un'adeguata comprensione di come funzionano le soluzioni e si producono i loro risultati. La tecnologia è *Explainable AI* fornisce trasparenza sui dati, sulle variabili e sui punti decisionali utilizzati per ottenere un risultato. Per contro, ricorre la realtà della cd. "scatola nera" (Black Box) in presenza di sistemi di IA/ML e di altre tecnologie innovative opache, non intuitive e che non forniscono informazioni adeguate sul processo decisionale e sulle previsioni/risultati

- 5. Le prime applicazioni di IA, big data e ML nelle esperienze di supervisione e di controllo delle autorità pubbliche: la vigilanza bancaria e il contrasto all'evasione fiscale.

Anticipando temi di interesse connessi all'applicazione dei *big data* alle attività di sorveglianza, una prima esperienza è maturata nella **vigilanza bancaria e finanziaria[61]**.

---

[59] Shapley Additive exPlainations (SHAP): tecnica di *explainabilty model-agnostic* basata sui cosiddetti *Shapley values*, usati nella teoria dei giochi per determinare quanto ogni giocatore (in un contesto collaborativo) abbia contribuito all'*output* prodotto complessivamente dal gruppo. Questi valori sintetizzano il contributo delle *feature* (proprietà individuali misurabili) del modello sull'output dello stesso,

[60] LIME (Local Interpretable Model-Agnostic Explainations): tecnica di *explainabilty model-agnostic* che intende spiegare il comportamento del modello mediante la variazione dei valori delle *feature* e l'osservazione degli impatti sugli *output* (in termini di contributo che ogni *feature* apporta alla predizione finale).

[61] Le applicazioni *Big Data* e *Machine Learning* sono destinate ad essere di forte ausilio per le attività sia del settore bancario/finanziario sia per le banche centrali. Sono in corso da anni discussioni su applicazioni rivolte a migliorare le attività tipiche delle banche centrali e a creare nuove metodologie per perseguire le loro funzioni istituzionali, come la politica monetaria, la supervisione bancaria e la supervisione sul sistema dei pagamenti. In tal senso è crescente il ruolo delle informazioni provenienti da fonti testuali e dalle immagini, nonché l'utilizzo di algoritmi di machine learning per combattere le attività finanziarie illecite e per aumentare l'accuratezza statistica di molti indicatori micro e

La Vigilanza della Banca d'Italia non utilizza ancora strumenti di IA, ma sta sperimentando applicazioni di ML, nel campo dell'**elaborazione del linguaggio naturale**[62].

In materia di **vigilanza prudenziale** (volta a preservare la sana e prudente gestione dei soggetti vigilati) di natura **ispettiva** sono in fase sperimentale strumenti che assistono gli ispettori nell'individuazione delle fattispecie analoghe e nei riferimenti normativi collegati, venendo in rilievo due classi di modelli, appartenenti alla "famiglia" dell'elaborazione del linguaggio naturale: l'analisi della similarità semantica tra oggetti testuali e la classificazione automatica delle informazioni che contengono. La digitalizzazione attraverso strumenti di IA potrebbe agevolare la redazione e la revisione dei rapporti ispettivi e migliorare la coerenza di trattamento. Nella **vigilanza a distanza** si analizzano molti dati quantitativi (dati di bilancio o metriche di rischio) ma possono rilevare per l'analisi anche dati qualitativi e non strutturati (testi che contengono anche numeri, tabelle e grafici, relazioni, verbali, minute, *audit reports* prodotte dagli organi aziendali e dai comitati interni degli intermediari vigilati) fornendo elementi utili a valutare il funzionamento e la qualità della *corporate governance* dell'impresa. Si sta perciò sperimentando l'applicazione di tecniche di *named entity recognition*, per individuare specifiche informazioni (entità) nei testi, e di *sentiment discovery*, per provare a ricavare un segnale informativo dalla documentazione aziendale. Un segnale composto di due "filamenti", includendo, da un lato, l'identificazione di questioni specifiche da approfondire, dall'altro, ove possibile, il tenore della dialettica interna agli organi aziendali. La Banca d'Italia ha condotto sperimentazioni anche nell'ambito della **vigilanza di tutela**, a cura del Dipartimento Tutela e educazione finanziaria. Queste hanno riguardato l'applicazione di tecniche di *text mining* e *machine learning* all'analisi di una particolare categoria di testi: gli esposti che la clientela bancaria presenta alla Banca d'Italia per segnalare pratiche potenzialmente scorrette da parte degli intermediari e presunte violazioni della disciplina sulla trasparenza[63]. L'IA può contribuire alla pre-analisi, quindi a migliorare l'efficienza del processo di esame di questa documentazione: può consentire di rintracciare temi comuni e questioni ricorrenti, in modo da accelerare l'individuazione di criticità che riguardano più parti del sistema. Nella maggior parte dei casi, sono utilizzati modelli di apprendimento supervisionato (*supervised learning*): richiedono un rilevante lavoro preparatorio sui dati, che richiede il contributo degli esperti di vigilanza. Siamo lontani dalla sostituzione degli analisti di vigilanza con algoritmi che possono essere d'ausilio ai primi in specifiche fasi dell'attività, per guadagnare in rapidità ed efficienza e migliorare la qualità dell'azione in termini di profondità di analisi e di uniformità di trattamento.

La Commissione europea, nel settembre 2020, con la sua comunicazione sulla *Digital finance strategy*, ha assunto l'impegno a sostenere l'adozione di strumenti *regtech* da parte dei soggetti vigilati e *suptech* (digitalizzazione dei processi di vigilanza) da parte dei supervisori, per rendere più efficiente anche il dialogo e lo scambio di informazioni tra le autorità e gli intermediari.

Sul fronte del **contrasto all'evasione fiscale** stanno maturando significative esperienze di applicazioni dell'IA.

Nell'**ordinamento francese** sin dal 2014 è stato istituito il sistema di *data mining* **CFVR** ("*Ciblage de la fraude et valorisation des requêtes*") che ricerca frodi ed errori nelle dichiarazioni dei redditi incrociando dati provenienti da banche dati pubbliche e private. In una logica di rafforzamento di questo tipo di analisi, l'art. 154 della Legge finanziaria francese per il 2020[64] ha autorizzato l'amministrazione fiscale e doganale, in via sperimentale e per la durata di tre anni, a raccogliere ed

macroeconomici. Tra le ragioni delle sfide provenienti da queste nuove metodologie centrale è la limitazione statistica dei Big Data dovuta alla mancata selezione di un campione, nonché i rischi relativi connessi con la privacy. Per alcuni contributi di riflessione cfr. https://www.bancaditalia.it/pubblicazioni/altri-atti-convegni/2019-bigdata/index.html.

[62] M. BEVILACQUA, L'INTELLIGENZA *artificiale e la vigilanza bancaria e finanziaria*, in F. FEDERICO F, MARCUCCI J, BEVILACQUA M, MARCHETTI DJ,. *Rapporto 2/2021 – L'impiego dell'intelligenza artificiale nell'attività di Banca d'Italia,* in BioLaw, 24 dicembre 2021.

[63] La Banca d'Italia ha ricevuto nel 2020 oltre 11 mila esposti su prodotti e servizi finanziari.

[64] LOI n. 2019-1479 del 28 dicembre 2019.

elaborare in via automatizzata le informazioni pubblicate dagli utenti sui *social network*. Oggetto del trattamento in questione sono tutti i contenuti che l'utente abbia deliberatamente divulgato sui propri profili *social*, con la sola esclusione di quelli accessibili dopo l'inserimento di una *password* o grazie ad un'apposita registrazione. I materiali così raccolti ed aggregati dall'algoritmo sono utilizzati dall'amministrazione per intercettare attività non dichiarate, verificare la corretta domiciliazione fiscale, ovvero per portare alla luce specifici illeciti quali, ad esempio, il traffico e la compravendita illegale di tabacco, alcolici o metalli preziosi che sono facilitati dall'uso di Internet. A tutela del cittadino viene escluso qualsiasi automatismo tra gli esiti del controllo algoritmico e il provvedimento amministrativo finale e vengono previste una serie di garanzie: i risultati dell'analisi automatizzata sono infatti trasmessi agli uffici amministrativi che hanno il compito di verificare ed eventualmente integrare quanto rilevato dall'algoritmo fiscale attraverso una ulteriore attività istruttoria ed essi rimangono competenti all'adozione della decisione amministrativa di avvio del procedimento o di archiviazione. La decisione di procedere o meno è quindi riservata al funzionario umano e rispetto ad essa l'indicatore informatico ha una mera funzione di supporto, finalizzata ad aprire una fase istruttoria che viene condotta con mezzi tradizionali. Sul funzionario grava quindi l'obbligo di distruggere entro cinque giorni dalla loro raccolta i dati non necessari, manifestamente irrilevanti o sensibili[65], oltre che un dovere di segretezza sui dati e le informazioni rilevati dal programma medesimo. L'ordinamento francese, all'avanguardia nell'automazione delle decisioni pubbliche, ha previsto espresse disposizioni a garanzia del cittadino poste all'interno del *Code des Rélations entre le public et l'administration* del 2016, così come modificato e integrato dal regolamento 2017-330 del 14 marzo 2017 (che ha introdotto gli artt. R. 311-3-1-1 e R. 311-3-1-2).

L'amministrazione è tenuta integrare il provvedimento con informazioni che consentano di conoscere le motivazioni e la logica seguita dall'algoritmo; garanzie non particolarmente rassicurante per il cittadino quando l'algoritmo stesso ha un ruolo di assistenza alla decisione e genera solo un avviso o una raccomandazione capace di proiettarsi in vario modo sull'acquisizione dei fatti rilevanti per la decisione e il provvedimento finale. La *Commission nationale de l'informatique et des libertés* (CNIL)[66], chiamata ad esprimersi al riguardo, pur facendo salvo l'algoritmo fiscale, ha manifestato alcune riserve[67]. In particolare, considerata l'importanza dei diritti fondamentali in gioco, quali la libertà di manifestazione del pensiero e la *privacy*, ha affermato la necessità di garantire non solo una **base legale** esplicita al trattamento di profilazione, ma anche il rispetto del **principio di proporzionalità** tra i dati analizzati e il fine perseguito dall'amministrazione, nel senso di limitare il trattamento solo ai dati strettamente necessari nonché pertinenti all'accertamento dell'illecito e alla gravità dell'infrazione contestata[68].

Un algoritmo con analoghe finalità è in uso da diversi anni nel **Regno Unito**, dove l'HMRC (*Her Majesty's Revenue and Customs*) ha introdotto nel 2012 un sistema denominato **Connect**, il quale utilizza varie tipologie di big data (in particolare informazioni provenienti dalle transazioni con carta di pagamento, registri catastali, conti bancari nel Regno Unito e all'estero, piattaforme online, compresi i mercati, i servizi P2P e i *social network*) per identificare i profili di potenziali evasori fiscali.

Il ricorso ad **applicazioni dell'IA**, seppure basate su **dati strutturati** e in una nozione "debole",

[65] Durante la pendenza di tale termine è vietato un loro utilizzo. Peraltro, solo i dati strettamente necessari a tale accertamento possono essere conservati, per un anno o, a seconda dei casi, fino alla conclusione del procedimento penale, tributario o doganale nell'ambito del quale sono utilizzati, mentre i dati che non sono idonei a contribuire all'accertamento delle infrazioni devono essere distrutti entro trenta giorni.

[66] Si tratta un'autorità amministrativa indipendente che svolge le funzioni di garante della *privacy* nel settore informatico in Francia

[67] CNIL, *Délibération* n. 2019-114 del 12 settembre 2019. L'autorità ha formulato diverse riserve volte a garantire uno stretto equilibrio tra l'obbiettivo di contrasto all'evasione e il rispetto delle libertà fondamentali, suggerendo un rafforzamento delle garanzie già previste dal legislatore. In particolare, la CNIL ha suggerito di individuare in modo specifico i dati rilevabili e le informazioni utilizzate, oltre a precisare il periodo di conservazione dei dati e di assicurare la distruzione di quelli non rilevanti

[68] G. AVANZINI, op. cit., *passim.*

si è sviluppato anche in Italia, specie nel settore dell'accertamento[69].

Si considerino, anzitutto, le **liquidazioni "automatizzate" delle dichiarazioni dei redditi ed iva** ai sensi degli artt. 36-*bis* d.p.r. n. 600 del 1973 e 54-*bis* d.p.r. n. 633 del 1972, per le quali vi è una precisa indicazione normativa ("avvalendosi di procedure automatizzate"). Tutte le dichiarazioni dei redditi e iva sono controllate al fine di liquidare imposte dovute e rimborsi spettanti[70]: gli uffici muovono da indicazioni precise[71], procedono entro confini prestabiliti a correzioni puntuali[72], agiscono in tempi brevi[73]. In tal modo vengono corretti errori che emergono *ictu oculi* dalle dichiarazioni con rideterminazioni di debiti e crediti emergenti dalle stesse; il che ha garantito risultati affidabili al punto che il legislatore ha esteso l'ambito di operatività di queste liquidazioni[74].

Altro impiego dell'IA nell'accertamento è dato dagli **indici sintetici di affidabilità fiscale (ISA)** di cui all'art. 9-*bis* d.l. 24 aprile 2017, n. 50[75]. Essi consentono di esprimere, su una scala da 1 a 10, il grado di affidabilità fiscale riconosciuto a ciascun contribuente, che, salve alcune esclusioni, eserciti attività di impresa, arti o professioni, per le quali sia approvato apposito decreto attuativo; dal 2018 sostituiscono gli studi di settore, dei quali condividono la funzione sostanziale di determinazione, tramite l'elaborazione dei dati contabili ed extracontabili forniti dal contribuente, di indici di normalità e coerenza della gestione aziendale o professionale del singolo contribuente. Tuttavia, mentre gli studi di settore venivano impiegati quali strumenti dell'attività di accertamento, imputandosi ai contribuenti i ricavi e compensi per essi determinati per il scostamento con quelli dichiarati utilizzandosi gli stessi valori quale presunzione aggiuntiva ai fini della determinazione del reddito[76], invece gli ISA dichiaratamente assolvono «al fine di favorire l'emersione spontanea delle basi imponibili e di stimolare l'assolvimento degli obblighi tributari da parte dei contribuenti e il rafforzamento della collaborazione tra questi e l'Amministrazione finanziaria»[77]. Ma ciò, naturalmente, non ne preclude l'utilizzo in sede di accertamento nei confronti dei contribuenti meno "affidabili", insieme ad altri elementi che concorrano in via presuntiva alla determinazione del reddito

---

[69] Per una panoramica cfr. A. GUIDARA, *Accertamento dei tributi e intelligenza artificiale: prime riflessioni per una visione di sistema*, in *Dir. Prat. trib.*, 2023, 2, 384.

[70] Non tutte le dichiarazioni sono soggette al diverso controllo formale di cui all'art. 36-*ter* d.p.r. n. 600 del 1973, selezionate "sulla base dei criteri selettivi fissati dal Ministro delle finanze, tenendo anche conto di specifiche analisi del rischio di evasione e delle capacità operative dei medesimi uffici".

[71] Il 2° comma di entrambi gli articoli (36-*bis* e 54-*bis* citt.) presenta la seguente formula: «sulla base dei dati e degli elementi direttamente desumibili dalle dichiarazioni presentate e di quelli in possesso dell'anagrafe tributaria».

[72] Cfr. art. 36-*bis*, comma 2, d.p.r. n. 600 del 1973

[73] Il 1° comma di entrambi gli articoli (36-*bis* e 54-*bis* citt.) stabilisce: «entro l'inizio del periodo di presentazione delle dichiarazioni relative all'anno successive» Anche se, come è noto, si tratta di un termine ordinatorio (cfr. art. 28, 1° comma, legge 27 dicembre 1997, n. 449) e il termine è quello di decadenza per la notifica della cartella di pagamento, previsto dall'art. 25, 1° comma, lett. *a*), d.p.r. n. 602 del 1973, il quale si applica anche all'iva in forza della previsione di cui all'art. 23 d.lgs. 26 febbraio 1999, n. 46.

[74] Sul tema cfr. A. ZUCCARELLO, *Algoritmi e automatismi nei controlli della dichiarazione: profili problematici*, in *Riv. tel. dir. trib*., 8 giugno 2022

[75] L'art. 9-*bis* è stato inserito, in sede di conversione del decreto, dalla l. 21 giugno 2017, n. 96 ed è stato poi più volte modificato, da ultimo dall'art. 24, 2° comma, d.l. 21 giugno 2022, n. 73, come convertito dalla l. 4 agosto 2022, n. 122.

[76] Per l'impiego degli studi di settore nell'attività di accertamento cfr. l'art. 10 della l. 8 maggio 1998, n. 146 e l'art. 62-*sexies* d.l. 30 agosto 1993, n. 331, come convertito dalla l. 29 ottobre 1993, n. 427. L'esclusione di ogni automatismo nell'applicazione degli studi di settori ai fini dell'accertamento e della loro valorizzazione come elemento di un più ampio ragionamento presuntivo (*id est* presunzioni semplici) è stato rimarcato dalle sentenze delle Sezioni Unite civili nn. 22635, 22636, 22637, 22638 del 18 dicembre 2009. Sul ruolo degli studi di settore, in una visione evolutiva, cfr. G. GIRELLI, *Gli studi di settore quale strumento "multifunzionale" tra dichiarazione, accertamento e processo tributario*, in *Riv. dir. trib*., 2012, I, 721 ss.; D. CONTE, *Il diritto di difesa del contribuente nell'ottica della Corte di Giustizia: il "passo del gambero" e il ritorno agli studi di settore come presunzione relativa?* in *Riv. dir. trib.*, 2019, IV, 10 ss. Sul passaggio dagli studi di settore agli ISA si vedano ad esempio, anche per i riferimenti: A. PURPURA, *Gli studi di settore tra compatibilità con la normativa comunitaria e l'introduzione degli indici di affidabilità fiscale*, *retro*, 2019, 764 ss.; A. COMELLI, *I principi di neutralità fiscale e proporzionalità ai fini della disciplina dell'iva europea e nazionale: dagli studi di settore agli indici sintetici di affidabilità fiscale*, *retro*, 2019, 1085 ss.

[77] Il fine è perseguito «anche con l'utilizzo di forme di comunicazione preventiva rispetto alle scadenze fiscali», nonché con «l'accesso al regime premiale di cui al comma 11»: così il 1° comma, dell'art. 9, d.l. n. 50 del 2017.

di imprenditori, artisti e professionisti, recuperandosi la funzione degli studi di settore, come ridisegnata dalla giurisprudenza più recente).

Un terzo impiego dell'IA, in via di sviluppo, è dato dalla **dichiarazione precompilata**. Introdotta in via sperimentale per le imposte sui redditi di lavoro dipendente e assimilati, se ne prevede la futura estensione agli altri redditi[78] e all'iva; viene predisposta dall'Agenzia delle entrate, utilizzando le informazioni disponibili nell'Anagrafe tributaria e i dati trasmessi da parte di soggetti terzi, tra i quali quelli contenuti nelle dichiarazioni e certificazioni rese dai sostituti d'imposta; è resa disponibile telematicamente e con congruo anticipo al contribuente, il quale può accettarla, apportarvi delle modifiche, presentare un'autonoma dichiarazione[79]. Pur non attendendo in sé all'accertamento, quest'ultimo ne risulta fortemente condizionato. Il contribuente è indotto ad accettare la dichiarazione precompilata, giacché gli vengono proposti dati dei terzi *prima facie* considerati (più) attendibili[80] e gli vengono offerti i vantaggi dell'esclusione dei controlli formali (e di dedicati "controlli preventivi")[81].

Tali applicazioni si rapportano con diverso livello di conformità ad alcune delle garanzie che devono assistere l'impiego dell'IA. Se la **parziarietà** è rispettata, non può dirsi altrettanto per la trasparenza degli esaminati impieghi dell'intelligenza artificiale. La parziarietà dell'impiego dell'IA trova fondamento nell'art. **3-*bis* della l. n. 241 del 1990** , il quale dedicato al più ampio "uso della telematica", e quindi anche dell'IA, ne ribadisce la funzione servente rispetto all'attività amministrativa e nell'art. 22 del Regolamento Generale sulla Protezione dei Dati (GDPR) del Parlamento europeo 27 aprile 2016, n. 2016/679/UE, il quale, dedicato in generale ai processi decisionali automatizzati relativi alle persone fisiche, e quindi anche a quelli amministrativi, ne prevede, salve eccezioni, la sostituibilità con processi diversi, quantomeno in alcune parti, su richiesta dell'interessato, fissando il diritto del predetto, salve eccezioni, a non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato (compresa la profilazione), che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona. L'impiego dell'intelligenza artificiale riguarda la fase istruttoria dei procedimenti, non solo di accertamento, all'interno della quale può assumere dimensioni variabili; in nessun caso, però, può sostituirsi alla volontà di provvedere, che connota la fase decisoria del procedimento e pone fine allo stesso. Del resto, la giurisprudenza amministrativa è chiara nell'affermare la collocazione procedimentale dell'algoritmo e l'imputazione dei risultati degli impieghi dello stesso al titolare del potere, anche ai fini delle correlate responsabilità[82(42)]. Anche a ipotizzare un'istruttoria totalmente algoritmica, che in procedimenti vincolati quali sono quelli di accertamento non può escludersi, fermi restando i limiti di cui si è detto, occorre pur sempre una volontà umana che verifichi e faccia propri i risultati dell'istruttoria.

Più problematica la compatibilità con il **principio di trasparenza**. Infatti, non sono conoscibili gli algoritmi che stanno alla base degli stessi e spesso, a fronte di formule normative generiche, non sono neppure conoscibili i dati dagli stessi impiegati. Mentre la giurisprudenza è netta nell'esigere una piena conoscenza di tali passaggi, affinché possa verificarsi che gli esiti del procedimento robotizzato siano conformi alle prescrizioni e alle finalità stabilite dalla legge o dalla stessa amministrazione, esigendo piena conoscenza relativa sia alla "formula tecnica" sia alle spiegazioni "che la traducano nella 'regola giuridica' ad essa sottesa e che la rendano leggibile e comprensibile, sia per i cittadini che per il giudice"[83]. Ciò vale per gli ISA e le dichiarazioni

[78] Il 2° comma, dell'art. 1, d.lgs. 21 novembre 2014, n. 175.
[79] Cfr. art. 1 d.lgs. n. 175 del 2014.
[80] Cfr. art. 4 d.lgs. n. 175 del 2014.
[81] Cfr. art. 5 d.lgs. n. 175 del 2014.
[82] Cfr. Consiglio di Stato, sent. n. 8472 del 2019.
[83] Cons. Stato, n. 2270 del 2019 cit.

precompilate: le norme non indicano i dati da impiegare[84] o lo fanno solo per alcuni di essi[85]. E soprattutto non sono conosciuti né conoscibili i criteri che portano all'elaborazione dei risultati: tant'è che i contribuenti effettuano spesso delle simulazioni, soprattutto per gli ISA, i cui esiti non sono decifrabili e in taluni casi l'amministrazione è espressamente tenuta a consentirle[86]. Ma ciò vale pure per le liquidazioni delle dichiarazioni: infatti, a dispetto di indicazioni precise e correzioni puntuali, pur sempre vi sono dati di partenza generici e all'amministrazione residuano margini di apprezzamento (ad esempio laddove autorizzano «sulla base dei dati e degli elementi... in possesso dell'anagrafe tributaria»)[87]; mentre i risultati delle liquidazioni (imposte dovute e rimborsi spettanti) dovrebbero essere frutto soltanto di criteri matematici[88].

L'obiettivo di perfezionare l'attività di **selezione dei soggetti** a maggior **rischio di evasione** rappresenta una priorità dell'Agenzia delle entrate. L'enorme mole di dati oggi disponibili, l'avvento della tecnologia, l'implementazione dell'IA non possono essere scissi dalla tutela dei dati personali e della riservatezza dei contribuenti. In proposito, va dato conto di un'ultima novità. **Ve.R.A**., acronimo di "Verifica dei rapporti finanziari", è il nuovo applicativo con cui l'amministrazione finanziaria mira ad effettuare le analisi di rischio e selezionare i contribuenti da sottoporre a controllo[89]. Gli strumenti offerti dall'intelligenza artificiale sono un'opportunità che l'Agenzia delle entrate ha illustrato nella circolare n. 21/E del 20 giugno 2022 contenente gli indirizzi operativi e le linee guida per il 2022 sulla prevenzione e contrasto all'evasione fiscale. Anche in tal caso, alla doverosa ricerca dei mezzi più efficaci per intercettare le violazioni fiscali si contrappongono tuttavia i limiti imposti dalla tutela della *privacy* dei contribuenti e l'esigenza quindi di contemperare le opposte ragioni, nei termini indicati nel Decreto del Ministero dell'Economia e delle Finanze del 28 giugno 2022.

Già nel 2021 l'Agenzia delle entrate aveva fatto riferimento[90] al progetto "*a data driven approach to tax evasion risk analysis in Italy*", un programma finanziato dall'Unione Europea che sfrutta le più moderne **tecnologie informatiche** (*network science*, intelligenza artificiale, *data visualization*), allo scopo di innovare i processi di valutazione del **rischio di *non-compliance*** dei contribuenti e che, nelle intenzioni della nostra Amministrazione finanziaria, permetterà di realizzare un nuovo sistema di supporto ai processi di individuazione dei **soggetti ad alto rischio** di frodi fiscali, attraverso l'introduzione di tecniche innovative di *risk analysis*. L'acquisizione ed utilizzazione di una grande mole di **dati ed informazioni** presenti nell'**Anagrafe tributaria** e nell'Archivio dei rapporti finanziari pone risvolti e criticità in tema di **tutela dei dati personali**: la garanzia dell'anonimato nella prima fase istruttoria, la possibilità di svelare il nominativo soltanto all'esito di incroci specifici, la precisazione dei tempi di conservazione dei dati, hanno persuaso il garante della *privacy* ad esprimere il parere favorevole, consentendo l'emanazione del Decreto del Ministro dell'Economia e delle Finanze del 28 giugno 2022 (in G.U. n. 152 del 1° luglio 2022) con il via libera al **nuovo applicativo per la Verifica dei Rapporti Finanziari** (**Ve.R.A.** l'acronimo). Un algoritmo con chiara funzione di deterrenza, che opera una deroga a numerose norme del codice della *privacy.*

L'elemento di maggiore novità della circolare n. 22/E/2022 dell'Agenzia delle entrate è

---

[84] Ciò accade in particolare per gli ISA, riferendosi genericamente il 1° comma dell'art. 9-*bis* d.l. n. 50 del 2017 cit. a «dati e informazioni relativi a più periodi d'imposta» e a «indicatori elementari tesi a verificare la normalità e la coerenza della gestione aziendale o professionale, anche con riferimento a diverse basi imponibili».

[85] Si pensi ai dati relativi a oneri e spese sostenute dai contribuenti che soggetti terzi trasmettono all'Agenzia delle entrate, di cui all'art. 3 d.lgs. n. 175 del 2014, rilevanti ai fini della dichiarazione precompilata

[86] Cfr. art. 9-*bis*, comma 5, d.l. n. 50 del 2017 cit.

[87] Cfr. ad esempio A. ZUCCARELLO, *Algoritmi e automatismi nei controlli della dichiarazione*, cit., ove pure si richiamano alcune pronunce della Cassazione a conferma dell'importanza dell'impiego dell'anagrafe tributaria e del *deficit* di trasparenza dei controlli automatizzati che contraddistinguono le liquidazioni

[88] In senso critico A. GUIDARA, *op. cit.* rileva che ampi margini di scelta e criteri eterogenei di elaborazione dei risultati, quali quelli più sofisticati dei sistemi di *machine learning* o di *black boxes AI*, esulano dai poteri organizzatori dell'amministrazione e richiedono precise indicazioni normative (ed opportune garanzie per i contribuenti), visto che per essi si addiviene a risultati, che sono di difficile comprensione e spiegazione, ma anche di difficile comparazione con quelli cui si addiviene per gli itinerari propri degli ordinari accertamenti analitici e deduttivi. L'A. rimarca come sia imposta la valorizzazione della partecipazione del contribuente nei procedimenti tributari: infatti il coinvolgimento del contribuente può condurre ad un ridimensionamento dei problemi avvertiti.

[89] M. CONIGLIARO, *Lotta all'evasione con l'intelligenza artificiale "Ve.r.a.***"**, in *Fisco*, 2022, 32-33, 3107.

[90] Cfr. Comunicato stampa del 4 marzo 2021.

rappresentato dalla "presentazione" di un nuovo strumento per il contrasto all'evasione: l'analisi del rischio basata sui **dati dell'Archivio dei rapporti** viene infatti potenziata mediante l'elaborazione, a cura del Settore Analisi del rischio e ricerche per la *tax compliance* della Divisione Contribuenti, di **nuove liste selettive** per l'attività di controllo, che sono **rese disponibili mediante l'applicativo Ve.R.A**. Il Decreto del Ministro dell'Economia e delle Finanze del 28 giugno 2022 (in G.U. n. 152 del 1° luglio 2022), visto l'art. 1, comma 682, della Legge 27 dicembre 2019, n. 160, prevede che l'Agenzia delle entrate si avvale delle tecnologie, delle elaborazioni e delle **interconnessioni** con le **altre banche dati** di cui dispone, allo scopo di individuare i criteri di rischio utili per far emergere le posizioni da sottoporre a controllo e incentivare l'adempimento spontaneo. L'entrata in vigore delle disposizioni contenute nel Decreto ministeriale costituisce l'avvio di una nuova stagione di **contrasto all'evasione** che risponde anche alle esigenze di attuazione del PNRR. Il punto di partenza di tali attività è costituito dall'**incrocio** fra i dati presenti nell'Archivio dei rapporti finanziari con le altre informazioni delle quali l'Amministrazione finanziaria è in possesso. Il punto debole del nuovo sistema di intercettazione dei soggetti a maggior rischio fiscale è costituito dalla possibilità che le elaborazioni effettuate tramite gli applicativi informatici subiscano delle intrusioni o delle alterazioni indesiderate[91]. Con il nuovo applicativo, l'Agenzia delle entrate potrà effettuare delle **elaborazioni finalizzate a far emergere le posizioni di contribuenti da sottoporre a controllo.** L'Agenzia delle entrate, con il Provvedimento direttoriale n. 176227/2022 del 23 maggio 2022, ha ampliato le **informazioni** che gli **istituti di credito devono trasmettere nel suddetto Archivio**, inserendo ad esempio le criptovalute, ed ha introdotto nuove funzionalità e modalità operative finalizzate alla predisposizione di un terreno il più fertile possibile ai fini dell'utilizzo delle nuove tecnologie selettive in chiave antievasione sulla base delle previsioni contenute nella Legge di bilancio 2020. Un vero e proprio "patrimonio informativo" di proprietà dell'Agenzia delle entrate nel quale sono presenti informazioni relative a ciascun codice fiscale: si va dalle dichiarazioni fiscali al patrimonio immobiliare e mobiliare dei contribuenti italiani, al loro tenore di vita sulla base delle spese annualmente sostenute, fino alle informazioni di natura contabile per i titolari di posizione IVA.

Come si accennava in premessa, non sono mancati i dubbi sul rispetto delle **norme sulla *privacy***.

A tal proposito, il Decreto ministeriale considera l'esigenza: - di fornire all'interessato le informazioni ai sensi dell'art. 14 ("Informazioni da fornire qualora i dati personali non siano stati ottenuti presso l'interessato"); - di consentire l'esercizio dei diritti di cui agli artt. 15 ("Diritto di accesso dell'interessato"), 17 ("Diritto alla cancellazione ("diritto all'oblio"), 18 ("Diritto di limitazione di trattamento") e 21 ("Diritto di opposizione") del Regolamento UE 2016/679; - di non arrecare un pregiudizio effettivo e concreto all'obiettivo di interesse pubblico di prevenzione e contrasto all'evasione fiscale. Viene altresì specificato, per taluni dati, il **diritto all'anonimato** del soggetto interessato. Il D.M. 28 giugno 2022, nel preambolo, richiama la **pseudonimizzazione**, termine con il quale si indica, ai sensi dell'art. 4, paragrafo 1, n. 5, del Regolamento UE 2016/679, il trattamento effettuato in modo tale che i dati personali non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e soggette a misure tecniche e organizzative intese a garantire che tali dati personali non siano attribuiti a una persona fisica identificata o identificabile. In pratica, si tratta di sostituire un nome con uno pseudonimo.

A tutela dei diritti e delle libertà degli interessati, l'Agenzia e la Guardia di Finanza sono chiamate quindi ad adottare le misure di sicurezza tecniche e organizzative idonee a garantire la **riservatezza**, l'**integrità**, la **disponibilità dei dati** e la **sicurezza dei sistemi**, nonché quelle necessarie ad assicurare che i dati utilizzati siano attuali, coerenti, completi, tracciabili e ripristinabili. In particolare, l'Agenzia delle entrate, anche per rafforzare le garanzie connesse al trattamento dei dati personali, deve effettuare le elaborazioni finalizzate a far emergere le posizioni da sottoporre a controllo su **dati preventivamente pseudonimizzati**, attraverso metodi di sostituzione o modifica delle informazioni anagrafiche ovvero tramite **perturbazioni delle variabili**, al fine di impedire, in presenza di dati finanziari, l'identificazione diretta degli interessati. L'affidabilità e

[91] Così A. BONGI, *Parte il nuovo algoritmo antievasione fiscale: analisi di rischio a prova di intrusioni*, in Quotidiano IPSOA, 2 luglio 2022.

l'accuratezza del modello di analisi e dei **criteri di rischio utilizzati** sono testati per fare in modo che all'esito delle analisi siano limitati i rischi di ingerenze nei confronti dei contribuenti che non presentano un rischio fiscale significativo e, comunque, siano limitati i rischi di erronea rappresentazione della capacità contributiva. Negli atti e nei provvedimenti indirizzati ai contribuenti vengono sempre illustrati il rischio fiscale identificato e i dati che sono stati utilizzati per la sua individuazione. Nel processo di formazione dei dataset di analisi e controllo è sempre garantito l'intervento umano. Gli indicatori di rischio desunti o derivati non vengono memorizzati in archivi o basi dati diversi dai data set di analisi e controllo e non sono utilizzati per finalità diverse dall'analisi del rischio di cui all'art. 1, commi da 682 a 686, della Legge 27 dicembre 2019, n. 160. Le attività istruttorie, nonché quelle di stimolo all'adempimento spontaneo, saranno in ogni caso condotte esclusivamente nei confronti dei soggetti che esercitano le funzioni di **rappresentanza legale**. Al fine di ridurre i rischi di accessi non autorizzati o non conformi alle finalità di trattamento, l'accesso agli strumenti informatici di trattamento è consentito ai soli soggetti specificatamente autorizzati, ai sensi dell'art. 29 del Regolamento e dell'art. 2-*quaterdecies* del D.Lgs. 30 giugno 2003, n. 196, deputati a svolgere le attività di misurazione della qualità dei dati e di analisi del rischio fiscale. Gli **elenchi elaborati a livello centrale**, mediante specifici criteri di rischio basati sull'utilizzo integrato delle informazioni comunicate dagli operatori finanziari all'Archivio dei rapporti finanziari e degli altri elementi presenti in Anagrafe tributaria, saranno utilizzati dalle Direzioni regionali e provinciali per indirizzare l'attività di controllo nei confronti delle posizioni a più elevato rischio di evasione, previa autonoma valutazione della **proficuità comparata**. Al fine di verificare la validità del modello di selezione, le **Direzioni provinciali**, coordinate dalle rispettive Direzioni regionali, avranno cura di comunicare gli esiti delle attività svolte su ciascuna posizione segnalata mediante la compilazione di un'apposita **scheda di *feedback***, resa disponibile all'interno dell'applicativo Ve.R.A.

- 6. Le strategie di contrasto per mezzo dell'IA del riciclaggio e del finanziamento al terrorismo.

- 6.1. Premesse.

Sempre più crescente è l'interesse verso la lotta al riciclaggio e al finanziamento del terrorismo mediante le nuove tecnologie, grazie anche alle nuove opportunità aperte dall'IA. Il riciclaggio di denaro e il finanziamento del terrorismo sono fenomeni con comune impronta transnazionale e assimilabili caratteristiche che esigono uniformi misure di contrasto, sulla base delle prescrizioni e delle linee di intervento provenienti dalla normativa comunitaria, dalle raccomandazioni del GAFI (Gruppo di Azione Finanziaria Internazionale) e dalle convenzioni e dalle risoluzioni delle Nazioni Unite. Le attività terroristiche, ad esempio, hanno bisogno di continue risorse (danaro, strutture logistiche, armi, documenti contraffatti, coperture e rifugi), sovente offerte dalla criminalità organizzata; le fonti privilegiate di finanziamento dei gruppi terroristici ricomprendono attività illecite quali i rapimenti a scopo di estorsione, il traffico di droga e le altre condotte illecite volte a reperire dei fondi, che, per essere riutilizzati nei canali leciti, devono essere necessariamente "ripuliti".

Ciò pone in evidenza come l'attività di riciclaggio sia intimamente connessa con il finanziamento al terrorismo, che può integrare reato presupposto del primo; d'altro canto, i fondi illecitamente acquisiti tramite traffici di droga o di armi o altre attività illecite delle associazioni terroristiche necessitano del reinvestimento nel circuito economico, con dissimulazione della provenienza. È questa la ragione criminologica che ha suggerito di impiegare le misure antiriciclaggio previste a livello nazionale e sovranazionale per il monitoraggio dei movimenti di denaro potenziando la trasparenza delle transazioni e delle attività finanziarie anche per il contrasto al finanziamento al terrorismo.

L'impiego dell'IA in tali ambiti discende dagli obblighi fissati dalle normative finalizzate a prevenire il riciclaggio[92] che sebbene non prescrivano un adempimento di essi servendosi di strumenti di IA, registrano l'esperienza di enti creditizi e intermediari di maggiori dimensioni impegnati nell'investire in sistemi algoritmici capaci di tracciare le transazioni e rilevare spostamenti di denaro

[92] Cfr. Regolamento UE 2015/847; direttiva UE 2015/849.

sospetti[93]. Tali normative potenziano la prevenzione del riciclaggio valorizzando secondo un approccio essenzialmente imperniato sul rischio, fondamentale per definire misure di prevenzione e controlli: tale metodologia *risk-based* presuppone che le banche e gli altri intermediari finanziari adottino un sistema in continua evoluzione, che documenti e aggiorni periodicamente la valutazione del rischio e la metta a disposizione delle Autorità, in Italia rappresentate dal Mef (Ministero dell'Economia e delle Finanze), dall'Uif (Unità Informazione Finanziaria), dalla Dia (Direzione Investigativa Antimafia) e dal Nspv (Nucleo Speciale Polizia Valutaria).

- 6.2. Le opportunità delle nuove tecnologie nel contrasto del riciclaggio e del finanziamento del terrorismo nell'analisi del GAFI.

La connessione tra il contrasto del riciclaggio e del finanziamento del terrorismo emerge nitidamente anche dall'analisi congiunta che di essi svolge il GAFI attraverso gli *International Standards*, nonché dalla pubblicazione delle "Opportunità e sfide delle nuove tecnologie per il contrasto del riciclaggio e del finanziamento del terrorismo"[94] e la successiva *Digital Strategy*[95]. L'innovazione responsabile è sostenuta da altre dichiarazioni internazionali, in particolare dalla risoluzione 2462 del 2019 del Consiglio di sicurezza delle Nazioni Unite, che invita tutti gli Stati a migliorare la tracciabilità e la trasparenza delle transazioni finanziarie, anche sfruttando appieno l'uso di tecnologie finanziarie e normative nuove ed emergenti per rafforzare l'inclusione finanziaria e contribuire all'attuazione efficace delle misure AML/CFT.

Il GAFI, in particolare, è impegnato a tenersi al passo delle tecnologie e dei modelli di *business* innovativi nel settore finanziario, dovendo garantire che gli *standard* globali rimangano aggiornati ed in grado di creare un settore finanziario "intelligente", governato da una regolamentazione capace di gestire i rischi e di promuovere l'innovazione responsabile[96]. Di conseguenza, per aumentare la consapevolezza dei progressi rilevanti nell'innovazione e nelle soluzioni digitali, nel rapporto del 2021 il GAFI ha esaminato le opportunità e le sfide delle nuove tecnologie per l'AML/CFT, senza trascurare l'analisi degli ostacoli al loro impiego e delle modalità per mitigarli.

Va premessa la centralità nel sistema antiriciclaggio degli adempimenti dell'**adeguata verifica della clientela** (*customer due diligence*, o CDD) e della **segnalazione di operazioni sospette;** inoltre, va ricordato che l'**approccio basato sul rischio** dovrebbe essere la pietra angolare di un efficace sistema AML/CFT, in quanto essenziale per una corretta gestione dei rischi. Pur a fronte delle linee guida del GAFI, la revisione strategica del quarto ciclo di valutazioni reciproche ha rilevato che molte giurisdizioni continuano ad applicare sistemi in gran parte basati su regole; parimenti, il settore privato continua a faticare ad adottare l'approccio basato sul rischio, preferendo un approccio costoso e difensivo all'AML/CFT. Una solida conoscenza e consapevolezza dei rischi, che consenta di mitigarli ed affrontarli in modo proporzionato, è fondamentale per l'efficace attuazione degli standard GAFI[97] e maggiore è la capacità di raccogliere ed elaborare dati, nonché di condividerli tra le parti

---

[93] M. PIZZOLLI, *Obblighi degli intermediari finanziari e intelligenza artificiale. Novità fiscali,* 2018 (7-8). pp. 348-356.

[94] FATF, *Opportunities and Challenges of New Technologies for AMLCFT,* Parigi, luglio 2021 in *https://www.fatf-gafi.org/publications/fatfrecommendations/documents/opportunities-challenges-newtechnologies-aml-cft.html*.

[95] FATF, *AML/CFT Digital Strategy for Law Enforcement Authorities,* in *https://www.fatf-gafi.org/*en*/publications/Digitaltransformation/Digital-transformation-law-enforcement.html,* 8 giugno 2022.

[96] Gli *standard* GAFI sono uno strumento dinamico che si evolve in risposta alle mutevoli minacce, vulnerabilità e rischi globali legati al riciclaggio di denaro e al finanziamento del terrorismo (ML/TF), e alle sfide che si verificano nella loro attuazione.

[97] L'approccio tradizionale, basato su regole fisse, ha portato all'osservanza difensiva, piuttosto che all'applicazione di diverse misure di attenuazione a diversi livelli di rischio mentre la risposta delle autorità all'eccesso di segnalazioni rispetto alla carenza di segnalazioni ha ulteriormente contribuito ad azioni difensive. Sia il settore pubblico che quello privato potrebbero non avere fiducia nelle proprie valutazioni del rischio a causa della loro comprensione incompleta della realtà, della mancanza di informazioni e dati e della mancanza di risorse e strumenti per effettuare valutazione del rischio solide, aggiornate e complete.

interessate, più sicuri sono i vantaggi significativi che possono provenire da un più dinamico approccio basato sul rischio.

L'a**deguata verifica della clientela**, sostanziantesi nelle procedure di identificazione, verifica e monitoraggio di essa, ha aumentato la trasparenza delle transazioni e reso più difficile per i criminali l'uso improprio dei prodotti finanziari; essa rappresenta uno dei pilastri fondamentali del quadro AML/CFT, che pure continua a presentare sfide in termini di attuazione ed efficacia.

Tra le **tecnologie innovative** il GAFI annovera: *i*) le soluzioni di **identità digitale**[98], che possono consentire l'identificazione/verifica dei clienti a distanza e l'aggiornamento delle informazioni, utili per migliorare l'autenticazione dei clienti per un accesso più sicuro ai conti e rafforzare l'identificazione e l'autenticazione quando l'adeguata verifica in fase di instaurazione (*onboarding)* e le transazioni vengono condotte di persona, promuovendo l'inclusione finanziaria e combattendo il riciclaggio di denaro, la frode, il finanziamento del terrorismo e altre attività finanziarie illecite[99]; *ii*) l'**elaborazione del linguaggio naturale**[100], che può supportare un'analisi più accurata, flessibile e tempestiva delle informazioni sui clienti e ridurre dati imprecisi, informazioni errate e consentire abbinamenti e ricerche più efficienti di dati aggiuntivi, favorendo valutazione del rischio più accurate, un migliore processo decisionale e riducendo i casi di esclusione finanziaria involontaria; *iii*) le soluzioni basate sulla tecnologia dell'AI e dell'apprendimento automatico (*machine learning* o ML) applicate ai ***big data***[101] possono rafforzare il **monitoraggio** e la segnalazione continua delle **transazioni sospette**, consentendone l'analisi automatica, distinguendole dalle normali attività, in tempo reale, riducendo al tempo stesso la necessità di una revisione umana iniziale in prima linea; *iv*) l'*Application Programming Interface* (API) e la *Distributed Ledger Technology* (DLT), la standardizzazione dei dati e delle normative in maniera da essere resi leggibili dalle macchine può aiutare le entità regolamentate[102] a riferire in modo più efficiente alle autorità di vigilanza e ad altre autorità competenti, consentendo avvisi, segnalazioni di *follow-up* e altre comunicazioni da parte di supervisori, forze dell'ordine o altre autorità alle entità regolamentate e ai loro clienti, nonché comunicazioni tra entità regolamentate e tra loro e i loro clienti; l'applicazione di analisi più avanzate da parte delle autorità di regolamentazione può anche rafforzare l'esame e la supervisione, anche fornendo potenzialmente un *feedback* più accurato e immediato.

---

[98] I sistemi o le soluzioni di identità digitale sono sistemi di identità o prodotti e servizi che eseguono il processo di identificazione/verifica dell'identità di una persona (fisica o giuridica), legando l'identità provata a una credenziale digitale e utilizzano le credenziali digitali e altri fattori di autenticazione per stabilire (confermare) che una persona che rivendica l'identità è la persona sottoposta a prova di identità, ovvero, è chi la persona afferma di essere.

[99] Tra i vantaggi dell'identità digitale per l'inclusione finanziaria per le entità regolamentate, si annoverano: i) la riduzione dei costi (l'ID può supportare processi più economici e sofisticati per l'*onboarding* dei clienti; unita alle maggiori possibilità di accesso ai servizi finanziari tramite dispositivi mobili e smartphone, la tecnologia può cambiare radicalmente il modo in cui i consumatori possono accedere ai servizi finanziari; Processi CDD più economici e automatizzati che consentono set di dati e fonti più ampi possono consentire ai clienti privi di referenze di credito di accedere a servizi finanziari o servizi di intermediazione automatizzati – e rendere tali servizi più convenienti); ii) la portabilità e interoperabilità(i sistemi possono essere utilizzati tra più istituzioni o transazioni riducendo l'onere della verifica a una sola istanza di *onboarding*, offrendo particolari vantaggi se la verifica iniziale è guidata dal governo); iii) ridurre l'errore umano, atteso che l'automazione della raccolta e della corrispondenza dei dati consente di prendere in considerazione molti più punti dati in un arco di tempo più breve di quanto sarebbe possibile eseguire manualmente); tra i vantaggi per gli individui sono ipotizzabili: (i) una migliore esperienza del cliente, riducendo il peso dell'identificazione personale e, ad esempio, la necessità di trasportare e inviare più documenti in formato fisico; ii) l'utilizzo multiplo dell'identità verificata, semplificando le operazioni quotidiane e offrono maggiore efficienza nelle interazioni con i fornitori di servizi e le autorità.

[100] La PNL è un ramo dell'IA che consente ai computer di comprendere, interpretare e manipolare il linguaggio umano e agli esseri umani di parlare con le macchine.

[101] Il *Financial Stability Board* definisce i big data come "l'enorme volume di dati generato dal crescente utilizzo di strumenti digitali e sistemi informativi", comprensivi di dati sulle transazioni finanziarie e sui social media, di dati meccanici (ad esempio, Internet delle cose), dati di computer e telefoni cellulari (FSB, 2017[25]

[102] Ai fini del Rapporto GAFI, per "entità regolamentate" si intendono gli istituti finanziari, i fornitori di servizi di asset virtuali (VASP) e le imprese e professioni designate non finanziarie (DNFBP), come definiti negli standard GAFI.

L’applicazione dell’apprendimento automatico e di altri strumenti basati sull’IA consente un’analisi dei dati in tempo reale, rapida e più accurata possono automatizzare parzialmente o completamente il processo di analisi del rischio, consentendogli di tenere conto di un volume maggiore di dati e di identificare i rischi emergenti che non corrispondono ai profili già compresi. Tali strumenti, in particolare, possono anche offrire mezzi alternativi per identificare i rischi, agendo di fatto come un controllo semi-indipendente sulle conclusioni dell’analisi del rischio tradizionale, rassicurando gli attori sulla completezza e accuratezza delle loro valutazioni, aumentando il loro grado di fiducia nell’applicazione di misure basate sul rischio, consentendo di giustificare più comodamente l’uso di tali misure ai supervisori. Gli strumenti automatizzati di valutazione del rischio possono anche essere più facilmente verificabili da parte delle autorità di vigilanza e offrire una maggiore obiettività.

L’apprendimento automatico e l’elaborazione del linguaggio naturale come le funzionalità basate sull’IA offrono grandi vantaggi all’AML/CFT per le entità regolamentate e le autorità di vigilanza.

L’**apprendimento automatico**, oltre a facilitare la gestione del rischio, essendo imperniato sulla capacità di apprendere dai sistemi esistenti, riducendo la necessità di input manuali nel monitoraggio, contrae i falsi positivi e consente di identificare i casi complessi. Ad esempio, le applicazioni di apprendimento automatico sono utili per rilevare anomalie, per identificare e per eliminare informazioni duplicate nonché per migliorare la qualità e l’analisi dei dati. Ad esempio, gli algoritmi DL[103], eseguendo ripetutamente un compito, attraverso progressive e lievi modifiche migliorative del risultato, consentono di risolvere problemi complessi senza l’intervento umano. Il monitoraggio delle transazioni tramite l’IA ed in particolare gli strumenti di apprendimento automatico può consentire alle entità regolamentate di svolgere funzioni tradizionali con maggiore velocità, precisione ed efficienza (a condizione che la macchina sia adeguatamente e accuratamente addestrata), filtrando i casi che richiedono ulteriori indagini, unendosi a sistemi di monitoraggio più ampi che includono un elemento di analisi umana per allarmi specifici o aree a rischio più elevato. Più in dettaglio, tra i settori antiriciclaggio nei quali il ML può aggiungere valore sono annoverabili: (i) l’identificazione e la verifica dei clienti (nel contesto dell’*onboarding* remoto e dell’autenticazione, l’IA, inclusa la biometria e l’apprendimento automatico è possibile eseguire analisi delle microespressioni, controlli anti-spoofing, rilevamento di immagini false e analisi degli attributi del volto umano); ii) monitoraggio del rapporto commerciale e analisi comportamentale e transazionale; iii) algoritmi di ML non supervisionati (raggruppare i clienti in gruppi omogenei per comportamento, utili per controlli impostati sulla base di un approccio basato sul rischio, ad esempio, attraverso delle soglie di transazione, consentendo un monitoraggio su misura ed efficiente); iv) algoritmi di machine learning supervisionati (per un’analisi dei dati più rapida e in tempo reale in base ai requisiti AML/CFT); v) punteggio degli avvisi (per concentrarsi su modelli di attività, emettere notifiche sulla necessità di una maggiore due diligence); vi) identificazione e implementazione di aggiornamenti normativi (in particolare le tecniche di machine learning con elaborazione del linguaggio naturale possono scansionare e interpretare grandi volumi di dati normativi non strutturati su base continuativa per identificare, analizzare automaticamente e quindi selezionare i requisiti applicabili o implementare i requisiti normativi nuovi o rivisti in modo che le entità regolamentate possano conformarsi ai prodotti normativi pertinenti); vii) segnalazione automatizzata dei dati (ADR), con uso di modelli di segnalazione standardizzati che utilizzano applicazioni digitali automatizzate rendendo disponibili in massa alle autorità di vigilanza le entità regolamentate sottostanti i dati granulari.

[103] Il *Deep Learning* (DL) è un tipo avanzato di apprendimento automatico in cui reti neurali artificiali (algoritmi ispirati al cervello umano) con numerosi strati (profondi) apprendono da grandi quantità di dati in modi altamente autonomi.

L'elaborazione del linguaggio naturale (NLP[104]) e gli strumenti di corrispondenza *fuzzy*[105] consentono anche una riduzione più efficiente dei falsi positivi e negativi (ad esempio nei processi di *screening* delle sanzioni), superando i problemi di qualità con il collegamento tra gli elementi di informazioni. L'Unità di Informazione Finanziaria, in collaborazione con la Direzione Generale per la Vigilanza e la Regolazione Finanziaria della Banca d'Italia, ha sviluppato un'applicazione di logica fuzzy per costruire indicatori AML per gli intermediari finanziari non bancari, che consente di elaborare dati quantitativi (pagamenti transfrontalieri da/verso paesi a rischio più elevato) al fine di supportare la valutazione periodica del rischio AML/CFT di tali intermediari. La fonte è il database delle segnalazioni antiriciclaggio aggregate (S.AR.A. dall'acronimo italiano) e delle segnalazioni di Vigilanza. Ai fini della costruzione degli indicatori, gli intermediari finanziari non bancari vengono suddivisi in diverse classi a seconda della loro tipologia (es. entità regolamentate per gli investimenti, società di gestione del risparmio, istituti di pagamento e di moneta elettronica, erogatori di credito) e attività principale (es. fondi aperti, fondi chiusi, fondi, trasferimento di denaro, moneta elettronica e altri servizi di pagamento, ecc.)

La **tecnologia di contabilità distribuita** (*Distributed Ledger Technology* , o DLT)[106] può offrire i seguenti vantaggi AML: (i) migliorare la tracciabilità delle transazioni su base transfrontaliera e persino su scala globale, rendendo potenzialmente più semplice la verifica dell'identità; ii) accelerare il processo CDD, poiché i consumatori possono autenticarsi e essere automaticamente accettati o rifiutati attraverso contratti intelligenti che verificano i dati; *iii*) previa definizione delle garanzie normative, le transazioni possono essere gestite tramite un unico registro condiviso tra diverse istituzioni in diverse giurisdizioni o tramite registri interoperabili, con potenziamento del monitoraggio, riduzione dei costi per il settore privato, con pool di dati più accurato e basato sulla qualità (In Cina, ad esempio, la DLT viene utilizzata dagli istituti finanziari per condividere liste di controllo o segnali d'allarme sulla base dell'ambito di riservatezza consentito da questo sistema).

L'analisi offerta dalla pubblicazione del rapporto GAFI rimarca come le nuove tecnologie possano rendere le misure antiriciclaggio (AML) e di contrasto al finanziamento del terrorismo (CFT) più rapide, economiche ed efficaci, migliorando l'attuazione degli *standard* GAFI e garantire

[104] È il ramo dell'IA che consente ai computer di comprendere, interpretare e manipolare il linguaggio umano e all'uomo di parlare ai computer.

[105] Si tratta di una logica polivalente, ossia un'estensione della logica booleana, legata alla teoria degli insiemi sfocati. È la tecnica logica che prende in considerazione dati imprecisi o approssimativi e li elabora utilizzando più valori, in modo da produrre un output utilizzabili, per quanto imprecisi, simulando il processo decisionale umano per estrarre informazioni di maggiore utilità. Quale sottoinsieme dell'IA da uno spettro aperto e impreciso di dati (input impreciso) elabora più valori in modo da produrre un output che includa una gamma di possibilità intermedie tra "si" e "no" (ad esempio, certamente sì, forse sì, non posso dirlo, forse no, certamente no). I sistemi *Fuzzy Logic* producono *output* definiti in risposta a *input* incompleti, ambigui, distorti o imprecisi (*fuzzy*), simulando il processo decisionale umano più della logica sì/no convenzionale. La logica *fuzzy* può essere implementata nell'*hardware*, nel *software* o in una combinazione di entrambi. Nella logica *fuzzy* (o logica sfumata o logica sfocata) si può attribuire a ciascuna proposizione un grado di verità diverso da 0 e 1 e compreso tra di loro. Con "grado di verità" o "valore di appartenenza" si intende quanto è vera una proprietà, che può essere, oltre che vera (= a valore 1) o falsa (= a valore 0) come nella logica classica, anche parzialmente vera e parzialmente falsa. L'IA e la logica fuzzy sembrano coincidere, in quanto la logica sottostante alle reti neurali è *fuzzy*. Una rete neurale, infatti, prende una serie di *input* valutati, attribuisce loro paesi diversi in relazione agli altri e giunge a una decisione che normalmente ha anche un valore. In questo processo non c'è nulla di simile alle sequenze di decisioni o-o che caratterizzano la matematica non fuzzy, quasi tutta la programmazione dei computer e l'elettronica digitale. Negli anni '80, i ricercatori erano divisi sull'approccio più efficace all'apprendimento automatico: modelli di "buon senso" o reti neurali. Il primo approccio richiede alberi decisionali di grandi dimensioni e utilizza la logica binaria, adattandosi all'*hardware* su cui viene eseguito. I dispositivi fisici possono essere limitati alla logica binaria, ma l'IA può utilizzare il *software* per i suoi calcoli. Le reti neurali adottano questo approccio, che si traduce in modelli più accurati di situazioni complesse; le prime hanno trovato posto in una moltitudine di dispositivi elettronici (C. ELKAN, *The paradoxical success of fuzzy logic*, in *IEEE Expert*, vol. 9, n. 4, 1994, pp. 3–49).

[106] La DTL o *blockchain* è un tipo di protocollo tecnologico che consente l'accesso, la convalida e l'aggiornamento simultanei di un registro immutabile (record digitale) distribuito su più computer e in genere, su più entità o posizioni.

l'inclusione finanziaria. Se in via generale, competenze, metodi e processi innovativi, nonché modi innovativi di utilizzare processi consolidati basati sulla tecnologia, possono aiutare i regolatori, i supervisori e gli enti regolamentati a superare molte delle sfide AML/CFT, in particolare la **tecnologia** può facilitare la raccolta, l'elaborazione e l'analisi dei **dati** e aiutare gli attori a identificare e gestire i **rischi** di riciclaggio di denaro e finanziamento del terrorismo (ML/TF) in modo più efficace e quasi in tempo reale. Se le nuove tecnologie rendono i pagamenti e le transazioni più rapide, le stesse agevolano l'accuratezza dei sistemi di identificazione, il monitoraggio, la tenuta dei registri e condivisione delle informazioni tra autorità competenti ed entità regolamentate. Così, il maggiore utilizzo di soluzioni digitali AML/CFT basate sull'intelligenza artificiale (AI) e sui suoi diversi sottoinsiemi (apprendimento automatico, elaborazione del linguaggio naturale) può aiutare a migliorare l'identificazione dei rischi, il monitoraggio, la reazione e la comunicazione delle operazioni sospette.

A livello del **settore pubblico**, un migliore monitoraggio in tempo reale e uno scambio di informazioni con le controparti consentono una supervisione più informata rispetto agli enti regolamentati, contribuendo a migliorare la vigilanza. A livello del **settore privat**o, la tecnologia può migliorare la valutazione del rischio, le pratiche di *onboarding*, i rapporti con le autorità competenti, la verificabilità, la responsabilità e la buona *governance* generale, risparmiando sui costi. L'uso di nuove tecnologie può migliorare l'efficacia dell'attuazione degli standard GAFI basata sul rischio, garantendone la compatibilità con gli standard internazionali di protezione dei dati, in punto di privacy e di sicurezza informatica.

Se utilizzate in modo responsabile e proporzionale, le innovative tecnologie AML/CFT possono aiutare a identificare i rischi e concentrare gli sforzi, ma la revisione manuale e il contributo umano rimangono importanti; anche in un contesto normativo abilitante la tecnologia, infatti, è necessario fare affidamento sugli attori umani per identificare e valutare eventuali rischi residui presentati dalle nuove tecnologie e mettere in atto misure di mitigazione adeguate. La combinazione dell'efficienza e dell'accuratezza delle soluzioni digitali con le conoscenze e le capacità analitiche degli esperti umani produce sistemi più robusti in grado di rispondere efficacemente ai requisiti AML/CFT pur essendo pienamente verificabili e responsabili.

- 6.3. L'impiego dell'IA nell'AML: le entità regolamentate.

Quante alle **entità regolamentate** (anzitutto gli istituti finanziari), gli esperti[107] hanno raggruppato le opportunità di impiego dell'IA al settore in cinque aree principali: *i*) la *compliance*, *ii*) l'antiriciclaggio e la *fraud detection*, *iii*) la gestione del credito e del risparmio, *iv*) la *cybersecurity*, *v*) le decisioni in materia di *trading* e di investimenti[108].

Da diverso tempo la *compliance* antiriciclaggio si fonda sull'analisi e valutazione di una ricca mole di informazioni, ancora governata da incarichi di tipo manuale, poco efficaci per contrastare le attività di antiriciclaggio e inefficienti in rapporto con gli investimenti operati nel settore.

Il **GAFI** ha sottolineato l'importanza dell'analisi dei *big data* tramite l'IA, le soluzioni tecnologiche basate sul *machine learning e* le altre tecnologie *AI-based* che consentono di implementare la ***customer due diligence*** (CDD) e la valutazione del rischio nonché di rafforzare il sistema di **monitoraggio** e **segnalazione di operazioni sospette.**

L'impiego del *ML* nel monitoraggio delle transazioni e dei soggetti può diventare decisivo per la gestione dell'enorme quantità di dati disponibili. Attualmente, le banche e gli istituti finanziari

[107] Cfr J. TRUBY, R. BROWN, A DAHDAL, *Banking on Al: mandating* a *proactive approach to Al regulation in the financial sector,* in *Law and Financial Market Review,* vol. 14/2020; luglio 2020.
[108] S. D'AMICO, *Strategie di contrasto per mezzo dell'intelligenza artificiale al riciclaggio ed al finanziamento al terrorismo,* in *Iusitinere.it,* 6.3.2023.

impiegano l'IA per ridurre gli oneri gravanti sui revisori umani, fornendo un concreto ausilio ai tradizionali metodi di monitoraggio delle transazioni. In particolare, i sistemi di ML si sono dimostrati validi nell'identificare i falsi positivi generati dal sistema di monitoraggio, riducendoli del 20-30%[109].

Al momento, le tecniche di ML sono applicate principalmente nei seguenti settori dell'AML e del CTF, per perfezionare: *i*) l'individuazione di **anomalie e** delle operazioni sospette, sia "conosciute" sia corrispondenti a nuovi schemi sospetti "sconosciuti"[110]; *ii*) l'automatizzazione della **raccolta di dati,** atteso che il *natural language processing* (NLP) e *l'optical character recognition* (OCR) possono attingere significato dai dati esterni "non strutturati" e accrescere le conoscenze dei *team* di *compliance*)[111]*; iii)* la **segmentazione dei gruppi di utenti o della clientela**; le tecniche di apprendimento non supervisionato possono aiutare gli esperti di *compliance* ad individuare modelli comportamentali che la revisione manuale non capterebbe; i sistemi di *clustering-data mining*, in particolare, raccogliendo dati "non etichettati" a seconda delle loro affinità o diversità, possono consentire una segmentazione più resistente alle variazioni dei dati, limitando i difetti attuali delle soglie nei sistemi di monitoraggio (la determinazione di un valore economico oltre il quale si attivano determinati meccanismi di verifica è una delle principali cause di "falsi positivi" che preclude un controllo accurato delle operazioni che si mantengono "sotto-soglia"; mediante la creazione di *cluster* di utenti sulla base delle informazioni processate dagli algoritmi *unsupervised*, è possibile invece predisporre soglie differenti a seconda del tipo di gruppo di soggetti individuato[112]); iv) la "**prioritizzazione**" **delle segnalazioni e del punteggio di rischio di un cliente**, attraverso impiego d algoritmi, accordando precedenza agli *alerts* che hanno un maggior grado di probabilità di essere realmente operazioni di riciclaggio di denaro[113]: tramite i modelli di *supervised learning,* addestrati a scindere le transazioni sospette da quelle regolari, si attinge dal gigantesco bacino di informazioni costituito dai precedenti esempi di segnalazioni di operazioni sospette, per individuare velocemente ed efficacemente le anomalie già censite[114].

Tra gli effetti benefici di tali sistemi sono annoverabili: *i*) il miglioramento dell'effettività dei sistemi antiriciclaggio; *ii*) il perfezionamento della gestione del rischio; *iii*) la riduzione dei costi, per gli istituti finanziari[115] in quanto il monitoraggio, l'elaborazione e l'analisi di transazioni sospette o altre attività illecite in maniera automatizzata, emancipandosi dall'iniziale revisione da parte dell'operatore umano, riduce la quota dei "falsi positivi". I modelli tradizionali di AML, infatti, generano, secondo alcuni studi, tra il 90 ed il 95%[116] di falsi positivi nelle segnalazioni, mentre secondo altri fino al il 98%[117]; inoltre la "gestione manuale" degli *alerts* basata sull'impiego di modelli statici appesantisce il carico di lavoro per i *team* di *compliance*, distraendone le risorse dai casi più ad alto rischio e rendendo inefficace dì fatto il ruolo stesso degli intermediari finanziari nel sistema antiriciclaggio.

Gli algoritmi supervisionati hanno però lo svantaggio di una **visione retrospettiva**, dipendendo da dati di addestramento noti, circostanza che può limitare l'individuazione di una nuova operazione

---

[109] Cfr. *https://www.thescienceofwheremagazine.it/2021/12/07/ml-vs-ml-machine-learning-versus-money-laundering-lintelligenza-artificiale-e-il-contrasto-al-riciclaggio.*

[110] B. BOUKHEROUAA, G. SHABSIGH, K. AIAJMI, J. DEODORO, A. FARIAS, E. SLSKENDER, AT MIRESTEAN, R RAVIKUMAR, *Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance,* in *Imf Departmental Papers,* Vol. 2021, *lnternational Monetary Fund.*

[111] Ibidem.

[112]Cfr. *https://www.thescienceofwheremagazine.it/2021/12/07/ml-vs-ml-machine-learning-versus-money-laundering-lintelligenza-artificiale-e-il-contrasto-al-riciclaggio.*

[113] Ibidem.

[114] S. D'AMICO, *Strategie di contrasto per mezzo dell'intelligenza artificiale al riciclaggio ed al finanziamento al terrorismo*, in *Iusitinere.it*, 6.3.2023.

[115] FATF, *Opportunities and challenges of new technologìes for AML/CTF*, luglio 2021

[116] DELOITTE-UOB, The *case far using Al in combating money laundering & terrorist financing- A deep dive into the application of machine learning technology; novembre* 2018; SAARADEEY, S, D GHOSH, R RAY, S GANESAN and R RAJAGOPALAN, *Disrupting status qua in AML compliance, ORACLE White Paper,* Marzo 2019

[117] McKinsey&Company, *Transforming approaches to AML and financial crimes, 2019*.

sospetta, con il rischio di identificare solo situazioni già individuate come corrispondenti a sospetti e non ciò che potrà risultare tale all'esito di investigazioni della FIU. Sussiste dunque il pericolo di incorrere nei cd "**falsi negativi**": dato che l'algoritmo apprende tramite le etichette precedenti, il suo grado di avanzamento non potrà di fatto che seguire quello del *team* di *compliance* dell'istituto finanziario che aveva in precedenza identificato le transazioni sospette[118]. Dunque, l'individuazione di anomalie deve auspicabilmente coordinarsi con *l'unsupervised learning*, cercando di identificare nuovi modelli sconoscendo a monte quali informazioni siano "indizio" di un'effettiva ipotesi di riciclaggio; solo tramite tale metodologia, è possibile individuare altri schemi di ML, altrimenti troppo complessi per essere colti dall'essere umano.

Nel 2021 è stato svolto un interessante studio sull'uso di IA, *Big data* e altre soluzioni tecnologiche avanzate da parte dei soggetti obbligati nell'ambito dell'antiriciclaggio e del contrasto al finanziamento del terrorismo[119]; l'indagine ha coinvolto un campione di 43 soggetti obbligati – banche, assicurazioni, altre istituzioni finanziarie, società di *gaming* – corrispondenti al 46% del totale attivo del settore finanziario nonché dei giochi e delle scommesse in Italia. Oltre ad esaminare il livello di impiego delle soluzioni tecnologiche avanzate da parte di questi soggetti, gli ambiti di applicazione nei processi AML/CFT, i benefici e i rischi derivanti dalla loro adozione, lo studio ha esaminato gli ostacoli che si frappongono ad un utilizzo delle "macchine intelligenti" su larga scala. Dalle conclusioni del rapporto dell'indagine emerge che: i) la diffusione delle soluzioni tecnologiche avanzate in ambito AML/CFT è ancora limitata, essendo adottate dal 53% dei rispondenti all'indagine, ma solo dal 39% dei soggetti di piccole e medie dimensioni (< 3.000 dipendenti); ii) tuttavia, l'84% dei rispondenti ha intenzione di investire in queste soluzioni nel prossimo futuro; iii) AI, analisi di *Big Data* e analisi testuale sono le soluzioni più utilizzate; in particolare, l'AI è utilizzata soprattutto in fase di monitoraggio delle transazioni (*transaction monitoring*), mentre il *big data analytics* osserva un impiego più trasversale in tutte le fasi dei processi AML/CFT; iv) tecnologie biometriche, sistemi basati su *blockchain*/*distributed ledger technologies* (DLT) e *cloud computing* sono le soluzioni meno impiegate; v) solo il 27% dei rispondenti si avvale di soluzioni interamente sviluppate *in-house*, mentre il 73% si affida, in maniera diversa, al supporto di *partner* esterni; rimane al momento limitata l'adozione di soluzioni *cloud-based; vi)* tra le fonti impiegate in fase di verifica AML/CFT, prevale l'uso di informazioni proprietarie (dati su operatività e anagrafiche della clientela) e le cosiddette 'liste compliance' (sanzioni, persone politicamente esposte/politici italiani locali, *enforcement);* i dati societari (da registri camerali e banche dati private) non appaiono utilizzati in maniera sistematica, soprattutto quelli con copertura globale, nonostante il carattere spesso transnazionale degli schemi di riciclaggio.

La verifica empirica ha il merito di dettagliare le aree AML di applicazione delle tecnologie avanzate, di seguito specificate: l'adeguata verifica in fase di acquisizione della clientela (*on-boarding)*; il monitoraggio delle transazioni (*transaction monitoring*); il monitoraggio continuo della clientela (*on-going monitoring*); la calibrazione dei modelli (es. definizione di soglie, *alert*); l'approfondimento investigativo tramite analisi di fonti aperte/dati non strutturati; la segnalazione di operazioni sospette alle autorità di sorveglianza; l'integrazione di dati/fonti aperte; la valutazione del rischio/segmentazione della clientela; la riduzione dei falsi positivi derivanti da applicativi AML/CFT in uso. Come detto, l'utilizzo di soluzioni evolute è maggiore in fase di adeguata verifica (***on-boarding***) e di monitoraggio delle transazioni (***transaction monitoring***) – ambiti nei quali, rispettivamente, il 94% e l'89% dei rispondenti che usa soluzioni avanzate dichiara di adottare almeno uno strumento - mentre per altre finalità è meno rilevante. La varietà di soluzioni adottate, poi, è diversa a seconda dell'ambito di utilizzo: se nell'*on-boarding* sono impiegate tutte le otto soluzioni indicate, in altri ambiti (es. monitoraggio delle transazioni) prevalgono alcune soluzioni

[118]Cfr. *https://www.thescienceofwheremagazine.it/2021/12/07/ml-vs-ml-machine-learning-versus-money-laundering-lintelligenza-artificiale-e-il-contrasto-al-riciclaggio.*

[119] M. NAZZARI, M. RICCARDI, *Next Generation AML: l'uso di big data e intelligenza artificiale nell'antiriciclaggio in Italia*, in Crime&tech, Università Cattolica del Sacro Cuore e SAS, Milano, 2021.

specifiche (come ad esempio l'AI e l'analisi di rete. Per chi impiega la AI, prevale l'utilizzo nel **monitoraggio delle transazioni e nell'approfondimento investigativo tramite analisi di fonti aperte/dati non strutturati.** Tra gli ambiti di applicazione dell'intelligenza artificiale nel settore AML osservati a livello internazionale emerge, soprattutto in ambito bancario, un utilizzo di modelli non supervisionati per identificare su larga scala anomalie comportamentali della clientela, che possano facilitare l'identificazione dei soggetti/transazioni su cui effettuare *due diligence* rafforzate e approfondimenti investigativi. Questi modelli processano informazioni soggettive, derivate dalle anagrafiche (es. età, classe di reddito, professione, origine/residenza), insieme ad informazioni sull'operatività e sul transato, per arrivare a una segmentazione della clientela in diversi *cluster* comportamentali e all'individuazione di quei soggetti che si discostano in maniera significativa dal comportamento medio.

Il ***Big data analytics*** è impiegato dal 56% degli utenti di soluzioni tecnologiche avanzate, mentre il 17% lo sta attualmente testando o ne sta attendendo l'implementazione. L'utilizzo di questa soluzione appare trasversale ai diversi ambiti di impiego, con una lieve prevalenza nel *transaction monitoring* e nella calibrazione dei modelli. L'**analisi testuale** risulta in uso da parte del 47% degli utenti di soluzioni avanzate, in fase di test o valutazione di adozione per un ulteriore 12%.

Alcune banche e istituti di pagamento a livello internazionale hanno iniziato a sperimentare l'utilizzo di strumenti di analisi testuale, più o meno evoluti, per estrarre valore da **documenti di testo non strutturat**i a fini AML/CFT. Un'applicazione particolarmente rilevante è legata all'**analisi delle causali** (o di altri campi testuali) collegate a **bonifici e ad altre operazioni** predisposte dalla clientela. L'analisi ha diversi obiettivi. In primo luogo, rilevare automaticamente la presenza di **parole chiave** potenzialmente utilizzate per occultare comportamenti fraudolenti o transazioni illecite (es. "*regalo*", "*donazione*", etc.) e altri elementi che il soggetto obbligato considera potenzialmente a rischio AML/CFT (es. nomi di beni di lusso, acronimi di compagini societarie estere, nomi di criptovalute). Dall'altra parte, l'analisi di questi campi testuali potrebbe **arricchire il patrimonio informativo sullo stile di vita e le abitudini di spesa della clientela**, integrare le informazioni anagrafiche già disponibili, dati transazionali e campi valore strutturati (es. importo, tipologia cliente, data). Questo approccio di *big data* consente, in ultima istanza, di avere a disposizione un set informativo più ampio sulla base del quale (i) migliorare la profilazione/segmentazione/*clustering* della clientela, (ii) rilevare anomalie, (iii) facilitare un'analisi 'qualitativa' della sproporzione tra abitudini di spesa e profilo reddituale.

L'**analisi di rete**, attualmente utilizzata dal 35% dei rispondenti che dichiarano di utilizzare soluzioni tecnologiche avanzate, e in fase di sviluppo o implementazione per un altro 24%, è prevalente nel *transaction monitoring* e nella gestione delle segnalazioni di operazioni sospette.

Al contrario delle soluzioni tecnologiche avanzate analizzate sopra, le **tecnologie biometriche**, il ***cloud computing*** e le soluzioni basate su ***blockchain***/DLT sono ancora scarsamente utilizzate. I realtà, potrebbero generare risparmi di tempo e aumento dell'efficienza negli accertamenti KYC, evitandone la ripetizione ogni volta che lo stesso cliente svolga un'operazione occasionale o instauri un rapporto continuativo con un altro soggetto obbligato della rete e potrebbe consentire un aggiornamento in tempo reale dei documenti contenuti nel *wallet* digitale, tenere traccia dei documenti caricati e delle diverse attività di *due diligence* svolte all'interno della rete di aderenti.

In più paesi sono state lanciate iniziative, in ambito bancario o di *partnership* pubblico-privato, con lo scopo di testare e adottare sistemi basati su *Blockchain* o altri approcci di *Distributed Ledger Technologies* (DLT), per semplificare i processi di acquisizione della clientela e, al tempo stesso, garantire l'efficienza degli adempimenti di *Know your customer* (KYC). In una di queste iniziative è stata realizzata una piattaforma basata su tecnologia DLT in cui diversi soggetti obbligati possono condividere informazioni sulla clientela, a fini di adeguata verifica. Il cliente carica le proprie informazioni anagrafiche su un *wallet* digitale che, in caso di autorizzazione da parte del cliente stesso, può essere condiviso dal soggetto che lo gestisce (*custodian*) con il richiedente (un altro

soggetto obbligato aderente alla rete).

Il ***cloud computing*** è utilizzato dal 20% dei rispondenti che hanno adottato soluzioni tecnologiche avanzate. Nella maggior parte dei casi (75%), questi sono dipendenti di **soggetti obbligati di piccole dimensioni (< 3.000 dipendenti)**. Il *cloud computing* permette, infatti, a questi soggetti di utilizzare determinate soluzioni avanzate, ottimizzandone però i costi e i requisiti tecnici necessari.

Le **tecnologie biometriche** vengono utilizzate dal 24% dei rispondenti che hanno adottato soluzioni tecnologiche avanzate. Il loro **impiego appare esclusivo nella fase di adeguata verifica in fase di *on-boarding***; se ci si poteva attendere un utilizzo maggiore di queste tecnologie, in una fase storica caratterizzata da un sostanziale aumento delle operazioni di *on-boarding* a distanza, anche a seguito della pandemia di COVID-19, il risultato pare correlato anche alle recenti disposizioni legislative introdotte per facilitare l'operatività a distanza hanno consentito un utilizzo più ampio delle misure di *strong authentication* e, correlativamente, hanno frenato l'impiego di tecnologie biometriche più avanzate. L'art. 27 del d.l. n.76/2020, convertito dalla l. n.120/2020 ha introdotto semplificazioni per i soggetti obbligati nell'adempimento degli obblighi di adeguata verifica dei clienti nei rapporti contrattuali che vengono avviati a distanza. Infatti, l'obbligo di identificazione del cliente e titolare effettivo si considera assolto anche senza la presenza fisica del cliente o esecutore quando: *a*) il cliente è in possesso di un'identità digitale con livello di garanzia "almeno significativo", rilasciata nell'ambito del Sistema pubblico per la gestione delle identità digitali e modalità di accesso ai servizi erogati in rete dalle pubbliche amministrazioni (c.d. SPID), come previsto dall'art. 64 del D.lgs. n. 82/2005 e dalla relativa normativa di attuazione; *b*) il cliente possiede un'identità digitale con livello di garanzia "almeno significativo" rilasciata nell'ambito di un regime di identificazione elettronica compreso nell'elenco pubblicato dalla Commissione europea a norma dell'art. 9 del Regolamento (UE) n. 910/2014 (c.d. Regolamento eIDAS); *c*) il cliente possiede un certificato per la generazione della firma elettronica avanzata (c.d. FEA); *d*) Il cliente è identificato per mezzo di procedure di identificazione elettronica sicure e regolamentate ovvero autorizzate o riconosciute dall'Agenzia per l'Italia digitale.

Quanto alle fonti ed alla natura **dei dati processati** dai soggetti obbligati in fase di verifica AML/CFT, è emerso che le informazioni che risultano più frequentemente impiegate sono: *i*) quelle 'proprietarie' (raccolte direttamente o disponibili all'interno del perimetro organizzativo del soggetto obbligato), consistenti nei dati relativi a **transazioni** e alle **anagrafiche della clientela**; *ii*) le informazioni provenienti dalle cd. "liste *compliance*", riferite alla sussistenza di **sanzioni**, a precedenti di natura giudiziaria ('***enforcement***'), alla presenza di **persone politicamente esposte** o **amministratori/politici locali** (liste '**PEP**' e '**PIL**'). Al contrario risultano ridotti: - *iii*) l'uso di **dati societari** (di origine camerale o derivanti da banche dati private) risulta più limitato, soprattutto per quelle provenienti da banche dati con copertura globale, utilizzate solo dal 24% del campione interpellato per tutti i soggetti sottoposti a verifica AML/CFT; *iv*) l'impiego di informazioni da '**whitelist antimafia**'[120] e da rapporti istituzionali come le relazioni semestrali ed annuali della Direzione Investigativa Antimafia, della Direzione Nazionale Antimafia e dell'Unità di Informazione Finanziaria della Banca d'Italia (UIF). Questo risultato potrebbe essere, da un lato, riflesso della **natura dei clienti** sottoposti a verifica (per la maggior parte persone fisiche residenti in Italia, per cui è più plausibile un'analisi dell'operatività e un incrocio con le 'liste', che un controllo sui registri camerali), ma anche di una **cultura del *risk assessment*** ancora troppo fondata quasi esclusivamente sulla rilevazione di sanzioni, precedenti giudiziari e PEP e non orientata – come richiesto dalla normativa di riferimento – a una visione olistica dei rischi e delle possibili anomalie che contraddistinguono i soggetti e le loro relazioni (societarie e non)

[120] Elenco dei fornitori, prestatori di servizi ed esecutori di lavori non soggetti a tentativo di infiltrazione mafiosa, operanti nei settori esposti maggiormente a rischio, così come individuato nell'art. 1 della Legge 190/2012 e modificato dalla Legge 40/2020.

con il contesto territoriale, settoriale, sociale, economico di appartenenza. Appare sorprendente, l'uso limitato di dati societari esteri, considerata la natura spesso transnazionale degli schemi di riciclaggio e di finanziamento del terrorismo segnalati dall'UIF e dalle autorità internazionali in materia, ad iniziare dalla *Financial Action Task Force* (FATF/GAFI).

- **6.4. L'impiego dell'IA nell'AML: le autorità.**

La sensibilità agli strumenti di *data analysis* si è diffusa anche tra le Autorità antiriciclaggio, sia nella funzione di Autorità di vigilanza che in quella di *Financial Intelligence Units* (FIU): tali organi stanno implementano il ricorso a strumenti che aumentano la loro capacità di rilevare reti di transazioni correlate, di identificare comportamenti anomali e, in generale, trasformare quantità significative di dati, strutturati e no, in informazioni utili a livello operativo. Questo approccio viene descritto con il termine ***SupTech***, *che* identifica l'uso da parte delle autorità finanziarie di strumenti avanzati di raccolta e analisi di dati, reso possibile dalle nuove tecnologie. Con riferimento all'impiego delle tecnologie nell'attività di supervisione, nel 2019, l'UIF ha pubblicato un quaderno che passa in rassegna gli strumenti avanzati di raccolta e analisi dei dati impiegati dalle autorità finanziarie con specifico riferimento al settore dell'antiriciclaggio[121]. Per fronteggiare gli incrementi del flusso segnaletico, nel 2011 è stato avviato l'impiego della Piattaforma per la raccolta e la gestione delle segnalazioni (RADAR); nel 2015 è stato costituito il c.d. *Data warehouse* che integra le basi dati utilizzate per le analisi; è stata messa a punto dapprima la gestione informatica degli scambi con gli Organi investigativi (2013) e poi con l'Autorità giudiziaria e le FIU estere attraverso un sistema dedicato, c.d. SAFE (2017). È inoltre in corso l'aggiornamento delle infrastrutture informatiche, la revisione del tracciato segnaletico, la riduzione delle aree di manualità nel processo di analisi mediante il ricorso a sistemi di intelligenza artificiale: nuove tecniche di risoluzione delle identità, analisi delle reti, aggiornamento e affinamento del sistema di rating delle SOS, criteri di selezione delle segnalazioni, per l'individuazione delle priorità e delle situazioni a più alto rischio.

Ad ottobre 2021, il GAFI, insieme all'*Egmont Group*[122] ha pubblicato il report «*Digital transformation of AML/CFT for operational agencies*», fornendo alle autorità finanziarie una guida operativa per la migliore individuazione dei *digital tools* per l'informazione finanziaria e l'investigazione di attività sospette, in base allo specifico obiettivo prefissato[123].

Ai fini limitati di rendere chiare le premesse giuridiche delle considerazioni che si verranno esponendo sulle possibilità applicazioni dell'IA e delle tecnologie avanzate nel settore AML, può essere utile ricordare che il sistema preventivo dell'antiriciclaggio delineato dal d.lgs. 21 novembre 2007 n. 231 è costruito su alcuni pilastri: *i*) la valutazione del rischio (art. 14-16); *ii*) l'adeguata verifica della clientela (art. 17-30); *iii*) la conservazione dei dati (artt. 31-34); *iv*) le segnalazioni di operazioni sospette (art. 35-41). A fianco, di questi dati aggregati che i soggetti obbligati sono tenuti a trasmettere all'UIF sulla propria operatività, al fine di consentire l'effettuazione di analisi mirate a far emergere eventuali fenomeni di riciclaggio o di finanziamento del terrorismo nell'ambito di determinate zone territoriali (art.33), le comunicazioni della pubblica amministrazione (art. 10, comma 4), le comunicazioni oggettive (art. 47) e i limiti all'utilizzo del contante (art. 49). L'Unità di Informazione Finanziaria per l'Italia riceve le segnalazioni di operazioni sospette dai soggetti obbligati, le analizza e le trasmette agli organi investigativi (la Guardia di finanza e la DIA), a loro volta chiamati ad analizzarli valutandone l'eventuale utilizzo per indagini in corso o per l'avvio di nuove indagini. L'adeguata verifica si articola (art 17 d.lgs. n. 231/2007) nelle seguenti attività: *a*) l'identificazione del cliente e la verifica della sua identità sulla base di documenti, dati o informazioni

[121] Cfr. Quaderno dell'antiriciclaggio, Collana Analisi e Studi n. 14, *Applicazioni suptech per l'antiriciclaggio*, ottobre 2019.
[122] Organismo globale delle Financial Intelligence Unit costituito nel 1995 per il supporto alle prassi operative e alla collaborazione internazionale.
[123] FATF- Egmont Group, *Digita/ transformation of AMUCFT for operational agencies, ottobre* 2021

ottenuti da una fonte affidabile e indipendente. Le medesime misure si attuano nei confronti dell'esecutore, anche in relazione alla verifica dell'esistenza e dell'ampiezza del potere di rappresentanza in forza del quale opera in nome e per conto del cliente; b) l'identificazione del titolare effettivo e la verifica della sua identità attraverso l'adozione di misure proporzionate al rischio ivi comprese, con specifico riferimento alla titolarità effettiva di persone giuridiche, trust e altri istituti e soggetti giuridici affini, le misure che consentano di ricostruire, con ragionevole attendibilità, l'assetto proprietario e di controllo del cliente; *c*) l'acquisizione e la valutazione di informazioni sullo scopo e sulla natura del rapporto continuativo o della prestazione professionale, per tali intendendosi, quelle relative all'instaurazione del rapporto, alle relazioni intercorrenti tra il cliente e l'esecutore, tra il cliente e il titolare effettivo e quelle relative all'attività lavorativa, salva la possibilità di acquisire, in funzione del rischio, ulteriori informazioni, ivi comprese quelle relative alla situazione economico-patrimoniale del cliente, acquisite o possedute in ragione dell'esercizio dell'attività. In presenza di un elevato rischio di riciclaggio e di finanziamento del terrorismo, i soggetti obbligati applicano la procedura di acquisizione e valutazione delle predette informazioni anche alle prestazioni o operazioni occasionali; *d*) il controllo costante del rapporto con il cliente, per tutta la sua durata, attraverso l'esame della complessiva operatività del cliente medesimo, la verifica e l'aggiornamento dei dati e delle informazioni acquisite nello svolgimento delle attività di cui alle lettere a), b) e c), anche riguardo, se necessaria in funzione del rischio, alla verifica della provenienza dei fondi e delle risorse nella disponibilità del cliente, sulla base di informazioni acquisite o possedute in ragione dell'esercizio dell'attività. Quanto alla segnalazione delle operazioni sospette, l'art. 35 d.lgs. stabilisce che "soggetti obbligati, prima di compiere l'operazione, inviano senza ritardo alla UIF, una segnalazione di operazione sospetta quando sanno, 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 o che comunque i fondi, indipendentemente dalla loro entità, provengano da attività criminosa. Il sospetto è desunto dalle caratteristiche, dall'entità, dalla natura delle operazioni, dal loro collegamento o frazionamento o da qualsivoglia altra circostanza conosciuta, in ragione delle funzioni esercitate, tenuto conto anche della capacità economica e dell'attività svolta dal soggetto cui è riferita, in base agli elementi acquisiti ai sensi del presente decreto. Il ricorso frequente o ingiustificato ad operazioni in contante, anche se non eccedenti la soglia di cui all'articolo 49 e, in particolare, il prelievo o il versamento in contante di importi non coerenti con il profilo di rischio del cliente, costituisce elemento di sospetto. La UIF, con le modalità di cui all'articolo 6, comma 4, lettera e), emana e aggiorna periodicamente indicatori di anomalia, al fine di agevolare l'individuazione delle operazioni sospette».

Dunque, la **SOS**: i) in termini **contenutistici**, consiste nella rappresentazione fornita da un soggetto obbligato di un'operatività anomala valutata come sospetta a seguito dell'apprezzamento professionale dei suoi elementi oggettivi e soggettivi; ii) sotto il **profilo formale**, si compone di dati strutturati relativi alle entità di interesse e di campi in formati testuale libero, destinati a riportare la descrizione dell'operatività e a rappresentare puntualmente la valutazione che ha indotto a ritenere sospetta (e non solo anomala) l'operatività di entrambe.

La **qualità** di entrambe le componenti (strutturata e libera) è fondamentale per i processi sottostanti alla valutazione da parte della UIF e successivamente degli organi investigativi. Secondo le analisi svolte dall'UIF sulle caratteristiche delle SOS pervenute dal 2020 ai nostri giorni, si registrano tendenze contrastanti. Ad un **aumento della complessità, non corrisponde un pari incremento della qualità.** Infatti, cresce il numero di soggetti, rapporti, operazioni e raccordi segnalati e così come quello dei contesti ampi e diversificati, piuttosto che delle operatività circoscritte, anche grazie all'applicazione di algoritmi più avanzati di rilevazione delle anomalie. Sennonché alla maggiore complessità non si associa sempre una qualità informativa più sviluppata. Lo rivelano le seguenti circostanze: *i*) la completezza dei campi descrittivi non sempre appare in linea con la complessità dei dati strutturati; ii) aumentano i casi di comunicazione di mera anomalia, senza comprensione, analisi e descrizione del fenomeno sospetto; *iii*) viene registrato l'invio di un numero

elevato di informazioni di dettaglio (nodi di una rete, transazioni, etc.) non messe in relazione tra loro e non siano spiegate, con aumenta del cd. "rumore di fondo" del sistema; iv) si identificano e vengono inviate SOS a fronte di mere anomalie, identificate con metodi automatici, finendo quasi per produrre comunicazioni oggettive (improprie) anziché SOS.

Il **prodotto informativo della UIF** si distingue tra quello offerto all'esito di analisi operative e quello offerto in funzione di analisi strategiche.

Le **analisi operative**, in particolare, sono volte a individuare specifiche situazioni o complessive **operatività di possibile riciclaggio** (attraverso le analisi delle SOS e delle comunicazioni oggettive, l'analisi dati aggregati, le collaborazioni con l'autorità giudiziaria, i raccordi con l'autorità di vigilanza, l'analisi di particolari flussi finanziari).

Le **analisi strategiche** sono orientate a individuare **categorie di rischio** anche prospettico verso cui indirizzare le attività della UIF o sulle quali richiamare l'attenzione di altre autorità (es. attuale i rischi connessi al PNRR).

Concentrando l'attenzione, anzitutto, sull'**analisi operativa** che riguarda le SOS, si possono distinguere **due fasi,** rispetto alle quali il possibile contributo degli strumenti di IA è chiaramente disomogeneo. Alla prima fase appartengono i sistemi automatici di **classificazione del rischio di riciclaggio** (*rating*) e gli strumenti di **classificazione automatica delle segnalazioni**. La seconda fase si compendia nella analisi finanziaria e nella stesura della relazione tecnica da parte degli analisti. L'uso della tecnologia è focalizzato principalmente sulla prima fase; in effetti, i principali interventi in termini di numeri trattati e di recuperi di efficienza attesi attengono alla revisione del sistema di classificazione del rischio (rating) e all'introduzione di strumenti di classificazione automatica delle segnalazioni[124].

In effetti presso l'UIF sono in elaborazione ed affinamento due progetti che attengono a questa tipologia di analisi[125]. Il primo *progetto* è basato su tecniche di *machine learning* e *deep learning*, finalizzato a modelli previsionali per **classificare in maniera autonoma, automatica** (senza l'intervento umano) le **segnalazioni di operazioni sospette**. Va considerato che la classificazione è fondamentale per l'UIF, considerando che a fronte di una dotazione di personale di circa 150-160 unità, dei quali 80-90 analisti, riceve quasi 150.000 segnalazioni all'anno; solo una corretta classificazione delle segnalazioni può consentire di concentrare l'analisi su quelle più rilevanti.

Un secondo progetto attiene a **indicatori di rischio di riciclaggio per le banche**. Utilizzando dati interni all'UIF, relativi a operazioni in contanti, bonifici verso paesi a rischio e altri tipi di operazioni intrinsecamente rischiose (traendo i dati dalle evidenze statistiche aggregate pubblicate nei bollettini) sono stati costruiti indicatori per ognuna delle circa 500-600 banche attive in Italia. Incrociando le statistiche di tutte le banche dati, l'UIF ha confrontato ciascuna banca in una certa provincia con se stessa nelle altre province e con tutte le altre banche nella stessa provincia, ad esempio per identificare gli istituti di credito che hanno una esposizione nelle operazioni in contanti o nei bonifici verso paesi a rischio o altre tipologie di strumenti di operazioni di pagamento intrinsecamente più rischiose e, dunque, quelli maggiormente esposti a questo rischio rispetto ai rispettivi 'pari' sul territorio e per tipologia di banca. L'indicatore, non sofisticato in termini statistici, è efficace per l'*intelligence* finanziaria e, in generale, per la supervisione bancaria contribuendo alla pianificazione ispettiva sia dell'UIF, sia della Vigilanza della Banca d'Italia, in relazione alla oculata scelta degli intermediari da sottoporre a controllo cartolare o ispettivo.

---

[124] C. CLEMENTE, *Il contrasto al riciclaggio e al terrorismo e la gestione dei congelamenti nella prospettiva della UIF*, relazione (inedita) al convegno «*L'utilizzo dell'intelligenza artificiale per il contrasto del crimine finanziario*», Torino, 13. Luglio 2022.

[125] D.J. MARCHETTI, *L'intelligence finanziaria: tre esempi di applicazione*, in F. FEDERICO, J. MARCUCCI, M. BEVILACQUA, DJ MARCHETTI, RAPPORTO *2/2021 – L'impiego dell'intelligenza artificiale nell'attività di Banca d'Italia,* in BioLaw, 24 dicembre 2021.

L'impiego della tecnologia nell'analisi operativa può avvenire secondo due modalità diversificate, ma complementari. Secondo un approccio *rule based* è possibile creare indicatori sintetici e algoritmi deterministici che, utilizzando dati strutturati e parole chiavi (*keywords*) presenti nei testi replicano il comportamento umano. Si tratta di strumenti introdotti da tempo, che contribuiscono, particolarmente, alle fasi iniziali dell'analisi finanziaria. A fronte del vantaggio della trasparenza delle logiche sottostanti ai risultati, vi è però il limite della necessità di predeterminare tutti i casi possibili. L'approccio ML, invece, utilizza le componenti testuali delle segnalazioni in ambiente dedicato (a tutela della riservatezza delle informazioni), con algoritmi che imparano da dati ma che scontano un approccio *black box*. I primi risultati, realizzati sperimentando su dati pregressi, usando insiemi ridotti di segnalazioni e di fenomeni, registrano livelli alti di sensibilità e accuratezza rispetto ai fenomeni più diffusi. L'evoluzione attesa è nel senso in una integrazione tra i due approcci.

Di assoluto interesse nel contesto dell'uso di strumenti di IA nell'analisi strategica è il progetto di ML volto ad individuare un indicatore di rischio di infiltrazione mafiosa per le imprese[126], basato sui dati di bilancio[127]. L'analisi è stata avviata a partire da un campione di circa 200 imprese, infiltrate dalla criminalità organizzata e oggetto di sequestro giudiziario; il campione, basato sui dati di ordinanze cautelari e altri provvedimenti giudiziari, è stato fornito dal ROS dei Carabinieri. I ricercatori dell'UIF hanno analizzato i bilanci di tali imprese con gli strumenti dell'econometria, confrontandoli con un campione rappresentativo dell'universo delle imprese italiane, in modo da individuare le caratteristiche di bilancio 'tipiche' delle imprese infiltrate. È stato quindi ottenuto un **indicatore di «somiglianza»,** alimentato con i dati pubblici del bilancio di un'impresa e che restituisce il grado di consonanza tra il bilancio dell'impresa e quello di un'impresa infiltrata operante nello stesso settore nella stessa provincia. Dalla prima fase, incentrata sull'econometria classica, si è passati all'utilizzo di tecniche di ML per costruire un indicatore del rischio di infiltrazione della criminalità organizzata nelle aziende sul territorio nazionale, sulla base dei dati di bilancio e di altri dati pubblici (localizzazione, settore economico, etc.). Per il periodo dal 2010-2021, è stato utilizzato un set formativo relativo a 16.074 imprese, con dati di bilancio, inclusivo di aziende sequestrate per reati di mafia e aziende con titolari o amministratori arrestati o indagati per mafia crimini correlati e si è esteso il campione su 1.832.848 imprese con dati di bilancio, che abbracciano tutti i settori e le regioni d'Italia. La **mappatura delle imprese potenzialmente prossime a contesti di criminalità organizzata è stata** ottenuta incrociando i dati Infocamere di tutte le imprese operanti in Italia (14 milioni di imprese e soggetti collegati) con i dati RADAR-DNA disponibili presso l'UIF[128]. Si è così attribuito un punteggio di rischio a livello aziendale da 0 a 1 (che può essere interpretato come probabilità di infiltrazione mafiosa). Il progetto ha consentito di pervenire al calcolo del punteggio di rischio su 931.163 aziende, per le quali vi era disponibilità di registri completi nel 2021. Il modello è stato segnalato per la capacità di riconoscere (sensibilità) l'81% delle imprese infiltrate presenti nei dati usati; inoltre, quando il modello segnala una determinata impresa sia infiltrata, l'indicazione predittiva è corretta (accurata) nell'87% dei casi[129].

Tale indicatore può essere utilizzato a più fini. Dalla prioritizzazione dell'analisi delle operazioni sospette presso l'UIF, agli indicatori aggregati di rischio per la ricerca strategica presso l'UIF, dall'individuazione degli intermediari con maggiore esposizione al rischio, sino agli impieghi investigativi da parte delle forze di polizia (ad esempio, screening massiccio di aziende che richiedono fondi pubblici). Può concludersi dunque che il forte aumento del numero e della quantità di dati e informazioni per finalità antiriciclaggio rendono l'utilizzo dell'IA uno strumento molto utile, sia il settore privato che le autorità.

---

[126] D.J. MARCHETTI, ult. op. cit.

[127] Questo indicatore è funzionale alle attività istituzionali dell'UIF perché fornisce un ulteriore strumento per valutare le operazioni sospette a carico di una certa impresa, arricchendo lo strumentario a disposizione degli analisti. Il sistema è stato realizzato *in house* e vi è stata una prima validazione, sia interna sia in collaborazione con organi investigativi.

[128] C. CLEMENTE, ult. op. cit.

[129] C. CLEMENTE, ult. op. cit.

Il cambiamento tecnologico tocca nel profondo i processi di lavoro degli analisti e le attività di monitoraggio strategico. Perché l'utilizzo dell'IA e dell'innovazione tecnologica in generale possa raggiungere i suoi obiettivi evitandone i rischi, occorre un cambiamento delle logiche organizzative che favorisca l'applicazione della *Data Science* all'antiriciclaggio. Le competenze accumulate nelle metodologie statistiche, econometriche e informatiche e nell'utilizzo dei metodi *big data* e *data analytics* devono valorizzare il ricchissimo patrimonio informativo a disposizione del sistema ponendosi al servizio degli analisti e delle loro professionalità.

Non va dimenticato però il rischio legato all'uso dell'IA nella segnalazione antiriciclaggio, potendo implicitamente incoraggiare un certo "disimpegno" da parte degli attori segnalanti, attraverso un'eccessiva dipendenza dall'IA stessa. Al contrario, l'esito degli algoritmi di IA deve essere sempre valutato attentamente intelligenza umana; inoltre, diversi crimini finanziari, non rilevati dall'IA, può essere individuato solo da un attento occhio umano: l'intelligenza artificiale è complementare e non sostituisce l'intelligenza umana[130].

- 6.5. L'impiego dell'IA nel contrasto del terrorismo.

- 6.5.1. Il momento definitorio, le tecniche di finanziamento e di contrasto del terrorismo.

La spinta verso la condivisione a livello sovranazionale della criminalizzazione delle diverse forme di terrorismo e del loro finanziamento ha segnato l'intera evoluzione convenzionale, nella quale robusta centralità e problematicità ha rivestito il **momento definitorio**. Le ulteriori linee di sviluppo della disciplina internazionale antiterrorismo sono rappresentate, *in primis*, dalla focalizzazione dell'attenzione sulle **transazioni economiche e finanziarie**, al fine di individuare le risorse impiegate per sovvenzionare le azioni terroristiche e i canali attraverso i quali esse pervengono alle organizzazioni; costante, ancora, è risultato il potenziamento della **collaborazione tra le Nazioni** per prevenire lo sfruttamento dei sistemi finanziari nazionali e internazionali a fini illeciti come il finanziamento delle attività terroristiche (c.d. fenomeno ***money dirtying***); da un lato, rendendo **omogenee le misure tecniche e legislative per contrastare tali fenomeni** e, dall'altro, sviluppando la **cooperazione giudiziaria**.

Nelle **tecniche di finanziamento del terrorismo** sono comunemente individuate **tre fasi**, simmetriche rispetto a quelle del riciclaggio[131]: la **raccolta** (*collection*), fase nella quale i fondi, di natura e origine sia lecita che illecita, raggiungono un collettore principale; **la trasmissione o l'occultamento** (*trasmission or dissimulation*), per celare le finalità ultime dei movimenti di capitale, utilizzando sistemi di pagamento "sotterranei", "paralleli" e alternativi al circuito bancario convenzionale (*underground* o *parallel banking systems*); **l'impiego** (*use*) del denaro o degli altri beni per il compimento di atti terroristici. Il **momento decettivo** caratterizza le operazioni del soggetto che investe denaro per finanziare il terrorismo non tanto (e non primariamente) rispetto all'origine delle risorse, quanto sul piano della loro **destinazione**, cosicché la corrispondente riprovazione involge prima che il meccanismo genetico delle risorse quello del loro **consumo**. Diversamente dal riciclaggio, infatti, le attività di finanziamento del terrorismo si prefiggono di **occultarne la destinazione**, più che la loro origine, e si realizzano con **somme anche esigue**, spesso movimentate e trasferite, come detto, mediante canali specifici (sistemi alternativi di trasferimento di fondi o attraverso l'interposizione di enti senza scopo di lucro), diversi da quelli finanziari ordinari. Non manca, però, una significativa **affinità di funzionamento tra i movimenti di capitali** diretti ad occultarne l'illecita provenienza e i flussi finanziari volti ad organizzare, favorire o porre in essere atti di terrorismo: in entrambi i casi, infatti, normalmente sono coinvolti una serie di **canali occulti**,

[130] D.J. MARCHETTI, *Financial crime detection: challenges and opportunities in using AI, a FIU perspective*, relazione inedita, 2023.
[131] RAZZANTE-RAMUNNO, *Riciclaggio e finanziamento al terrorismo di matrice islamica*, Filodiritto, 3.5.2007.

presenti anche in settori e centri finanziari caratterizzati da forte opacità; non di rado, poi, vengono realizzate **movimentazioni finanziarie dissimulate sotto operatività e ragioni economiche fittizie**. Per queste ragioni gli **strumenti dell'analisi finanziaria**, già utilizzati, a livello internazionale, **nella lotta al riciclaggio, sono stati estesi all'azione di contrasto del finanziamento del terrorismo internazionale**, pur nella consapevolezza che quest'ultimo reperisce le liquidità necessarie - anche - attraverso l'utilizzo di canali informali e lo sfruttamento dell'economia legale.

Nelle **tecniche di contrasto** del finanziamento del terrorismo, poi, la ricostruzione delle "tracce" dei capitali movimentati è condotta per individuare e **bloccare il finanziamento** dell'attività terroristica, prima ancora che per salvaguardare l'integrità del sistema economico contro forme di inquinamento; il settore finanziario, potendo garantire un livello di opacità delle operazioni superiore al normale (la c.d. asimmetria delle informazioni), si rivela fortemente appetibile per le organizzazioni terroristiche capaci di sfruttare le potenzialità offerte dall'integrazione globale dei mercati finanziari per trasferire capitali da un Paese all'altro senza essere identificati e senza lasciare tracce dell'operazione. Anche nell'azione di contrasto del finanziamento del terrorismo, dunque, è decisiva e irrinunciabile l'attenzione sui canali di trasferimento del denaro.

Dovendo rinviare ad altre riflessioni per **l'evoluzione internazionale** che ha accompagnato la normativa di contrasto del terrorismo[132], va qui ricordato che l'adozione da parte del **legislatore italiano** di adeguate misure di contrasto al **terrorismo internazionale** non è avvenuta attraverso un *corpus* normativo organico e autonomo, ma mediante una disciplina frammentaria in esito ad interventi, stratificati nel tempo, assunti con logiche emergenziali[133], con i quali sono state criminalizzate le diverse forme del terrorismo internazionale, in conformità alle indicazioni sovranazionali. Per contrastare il terrorismo internazionale, infatti, nel diritto repressivo e procedurale, sono state introdotte nuove norme penali, modificate le disposizioni di ordinamento penitenziario e processuali; sul piano preventivo e amministrativo, poi, sono state rafforzate le misure di prevenzione ai sensi del d.lgs. n. 159/2011 nonché introdotti nuovi sistemi di controllo finanziario e di congelamento delle risorse economiche ai sensi dei D.lgs. n. 231/2007 e 109/2007.

Inizialmente il legislatore nazionale ha preferito riproporre alcuni **strumenti repressivi e istituti già sperimentati** per fronteggiare altre manifestazioni criminali di carattere associativo come la mafia o il terrorismo di matrice politica interna [134], adattati alle nuove contingenze[135]. L'impianto del codice Rocco, in effetti, era particolarmente attrezzato nella tutela penale della personalità dello Stato e le tecniche di tutela codicistiche si sono prestate a contrastare il terrorismo interno degli anni Settanta. Un modello di progressione di tutela dalle condotte preparatorie sino all'attuazione del programma politico eversivo: dai delitti di istigazione/apologia/propaganda, alle fattispecie di accordo/associazione per finire con i delitti di attentato. Il modello garantiva l'anticipazione della tutela penale che, da strumento di lotta al "nemico politico" degli anni Trenta, si è venuto trasformando in strumento utile contro il nuovo nemico del "terrorista interno" degli anni Settanta del

[132] Per una possibile individuazione di *periodi* all'interno dell'operatività dello Stato Islamico, e per la ricostruzione dei *relativi fenomeni criminali*, cfr. M. ROMANELLI, *Criminalità organizzata e terrorismo: la circolazione dei modelli criminali e degli strumenti di contrasto*, in *Sistema penale*, 20 dicembre 2019; per l'evoluzione del quadro internazionale e nazionale cfr. F. DI VIZIO, *Il finanziamento del terrorismo: evoluzione della normativa penale e preventiva*, https://www.dirittogiustiziaecostituzione.it.

[133] DI BITONTO, *Terrorismo internazionale, procedura penale e diritti fondamentali in Italia*, in Cass. Pen., fasc. 3, 2012, pp. 1181 ss.

[134] VIGANÒ, *Il contrasto al terrorismo di matrice islamico-fondamentalistica: il diritto penale sostanziale*, in Terrorismo internazionale e diritto penale, a cura di de Maglie e Seminara, Cedam, 2007, p. 159.

[135] DI STASIO, *La lotta multilivello al terrorismo internazionale. Garanzie di sicurezza versus tutela dei diritti fondamentali*, Milano, 2010, p. 564. Sulla evoluzione della normativa interna dagli anni Settanta del XX secolo al 2005, in prospettiva penalistica, cfr. PASQUA, *Legislazione italiana antiterrorismo*, in BASSIOUNI (a cura di), La cooperazione internazionale per la prevenzione e la repressione della criminalità organizzata e del terrorismo, Milano, 2005, p. 405 ss.

secolo scorso[136]. La legislazione dell'emergenza ha segnalato l'esasperazione degli elementi di specialità del diritto penale politico: espansione dei reati associativi e dei delitti di attentato; potenziamento dei reati di opinione; forte accentuazione del profilo soggettivo, attraverso l'onnivora finalità di terrorismo o di eversione la quale, come elemento costitutivo o circostanza aggravante, è divenuta la chiave di accesso al sistema normativo di contrasto al terrorismo con disposizioni speciali in materia di diritto penale, diritto penitenziario, diritto processuale, misure di prevenzione. Il legislatore ha inasprito, inoltre, il trattamento sanzionatorio attraverso l'aumento dei limiti edittali di pena e l'introduzione di limiti al giudizio di bilanciamento delle circostanze (art. 1, l. 15/1980).

Lo sviluppo del **terrorismo internazionale** ha importato il passaggio dalla tutela progressiva del codice Rocco alla **tutela penale a "costellazione"**[137], attenta alla repressione delle **attività collaterali al fenomeno associativo**. Nella fenomenologia del terrorismo internazionale, infatti, la tradizionale **organizzazione**, asse dei reati associativi, è frequentemente disarticolata in una struttura a rete che rende più fluidi i contatti e le modalità di azione che si avvalgono dei c.d. **lupi solitari**[138] il cui collante è l'ideologia che fomenta la violenza. L'**espansione e l'anticipazione della tutela penale rispetto a condotte collaterali al fenomeno associativo** riflettono criminologicamente le modalità di azione delle associazioni terroristiche e di coloro che, per condivisione ideologica, programmano autonome condotte di supporto al programma [139]. Questo mutamento dell'indirizzo di politica criminale è in linea con la **disciplina sovranazionale** che ha imposto l'anticipazione dell'intervento penale, presente anche nell'ultima direttiva dell'Unione europea (2017/541) che ha sollecitato gli Stati membri a dare rilevanza penale ad ulteriori condotte (art. 8)[140]. Sul piano della tecnica di tutela, poi, hanno trovato ingresso fattispecie a tutela anticipata (artt. 270-*ter* ss. c.p.) connotate da tre elementi: *fattispecie* strutturalmente *sganciate dall'associazione con finalità di terrorismo* (applicabili fuori dai casi di concorso nel reato associativo, come indicato dalla clausola di riserva) con potenziamento anche dei reati di apologia e istigazione; valorizzazione del *dolo specifico* quale elemento strutturale che collega la condotta incriminata a ulteriori condotte con finalità di terrorismo anche molto distanti, temporalmente e spazialmente, dalla prima; parificazione del trattamento sanzionatorio delle fattispecie collaterali al fenomeno associativo alle condotte di partecipazione all'associazione. Al contempo, le nuove fattispecie a tutela anticipata sono funzionali al processo penale in chiave di *agevolazione probatoria*, sollevando il giudice dall'accertamento degli elementi della condotta di partecipazione o di concorso esterno che, in relazione alla criminalità organizzata di tipo mafioso, ha vissuto un progressivo affinamento giurisprudenziale con delimitazione della funzione incriminatrice dell'art. 110 c.p. ; sul piano cautelare, poi, tali figure criminose hanno permesso un ulteriore arretramento nell'intervento delle misure processuali.

Un **primo gruppo di interventi** *ad hoc* è stato realizzato solo dopo gli attentati dell'11 settembre 2001. Tre decreti-legge hanno attuato le misure contro *Al-Qaeda* e i Talebani (**DL 28 settembre 2001, n. 353** conv. in l. 27 novembre 2001, n. 415), introdotto misure volte a reprimere il finanziamento al terrorismo, istituendo il Comitato di sicurezza finanziario (**DL 12 ottobre 2001, n. 369** conv. in l. 14 dicembre 2001, n. 431) e operato modifiche al codice penale e al codice di procedura penale, rimodulando le norme esistenti per fronteggiare il terrorismo interno, in modo da renderle funzionali al contrasto del terrorismo internazionale (**DL 18 ottobre 2001, n. 374** conv. in l. 15

[136] PELISSERO, *La legislazione antiterrorismo. Il prototipo del diritto penale del nemico tra garanzie e rischi di espansione* in Rivista Italiana di Diritto e Procedura Penale, fasc. 2, 1/6/2020, p. 750.
[137] L'osservazione è di PELISSERO, *op. cit.*, p. 751.
[138] Cass. Pen. Sez. 1, 9 ottobre 2018, n. 51654.
[139] VIGANÒ, *Minaccia dei "lupi solitari" e risposte dell'ordinamento: alla ricerca di un delicato equilibrio tra diritto penale, misure di prevenzione e diritti fondamentali della persona*, in KOSTORIS - VIGANÒ (a cura di), Il nuovo pacchetto antiterrorismo, Torino, 2015.
[140] La **direttiva (UE) 2017/541** del Parlamento europeo e del Consiglio, del 15 marzo 2017, sulla lotta contro il terrorismo e che sostituisce la decisione quadro 2002/475/GAI del Consiglio e che modifica la decisione 2005/671/GAI del Consiglio (GU L 88 del 31.3.2017, quale recepita nel diritto nazionale, è **il punto di riferimento delle autorità nazionali competenti per la definizione dei reati di terrorismo**.

dicembre 2001, n. 231). In particolare, è stato introdotto il reato di *associazione con finalità di terrorismo internazionale* (art. 270-*bis* c.p.), legittimando allo svolgimento delle indagini gli uffici della Procura della Repubblica presso il Tribunale del capoluogo del distretto di Corte d'appello (art. 51, comma 3-*quater*, c.p.p.)[141] ed estendendo anche alle indagini relative al terrorismo internazionale alcune misure già previste in materia di contrasto alle organizzazioni criminali di stampo mafioso.

È prevista l'applicabilità delle disposizioni di cui all'art. 13 DL 13 maggio 1991, n. 152, convertito, con mod., dalla l. 12 luglio 1991, n. 203, nei limiti dell'art. 3 DL n. 374/2001[142], con possibilità di effettuare intercettazioni telefoniche, ambientali e di flussi informatici in presenza di sufficienti indizi di reato e di necessità delle intercettazioni per i delitti commessi per finalità di terrorismo; si apre all'ammissibilità delle intercettazioni preventive, su autorizzazione del p.m., escludendo valore probatorio ai risultati (art. 5 D.L. n. 374/2001[143]); viene introdotta la possibilità per gli appartenenti alle forze di polizia di svolgere attività sotto copertura in relazione al contrasto del terrorismo internazionale, l'applicazione delle misure di prevenzione personali e patrimoniali originariamente previste contro la mafia (art. 7 D.L. n. 374/2001) sono estese nei casi di commissione dei reati con finalità di terrorismo anche internazionale; viene autorizzato l'impiego del sistema della videoconferenza per l'esame e partecipazione a distanza degli imputati detenuti e dei collaboratori di giustizia (art. 8 DL. N. 374/2001).

L'espressa riconducibilità al terrorismo internazionale della fattispecie penale dell'art. 270-*bis* c.p., con i connessi adattamenti procedurali, è risultata decisiva per agevolare il contrasto di cellule a specifica connotazione etnico-religiosa le quali, utilizzando il territorio italiano come base logistica, sostenevano l'azione di organizzazioni terroristiche operanti in altri Paesi[144]. Per la Cassazione, infatti, l'originaria formulazione dell'art. 270-*bis* c.p. (introdotto *dall'art. 3 DL 15 dicembre 1979, n. 625, conv., con mod., nella l. 6 febbraio 1980, n. 15*), punendo associazioni clandestine terroristiche aventi come obiettivo la programmazione e l'esecuzione di aggressioni eversive contro lo Stato italiano (e non contro lo Stato estero) ne precludeva l'applicazione al terrorismo internazionale[145]; onde, prima della ricordata modifica dell'art. 270-*bis* c.p., la possibilità di perseguire in Italia gruppi organizzati criminali anche clandestini i quali, attraverso le attività svolte sul suolo italiano, avessero favorito la commissione di atti violenti contro l'ordinamento di Stati stranieri, era consentita nei ristretti limiti in cui fosse integrato il reato di associazione a delinquere (art. 416 c.p.) o altri reati strumentali a quell'obiettivo, senza l'impiego degli strumenti investigativi processuali e penali più idonei ad affrontare il fenomeno terroristico già collaudati in passato per la repressione del terrorismo interno (intercettazioni, custodia cautelare, trattamento penitenziario, ecc.)[146]. Inoltre, se ai fini del riconoscimento della sussistenza del reato di criminalità organizzata (di matrice eversiva o di stampo mafioso) la giurisprudenza consolidata esigeva il radicamento territoriale del gruppo, la stabilità logistica e la condivisione quasi totale tra i vari associati della conoscenza del programma e degli obiettivi a breve e medio termine dell'organizzazione di cui essi fanno parte[147], tali caratteristiche non erano riscontrabili

---

[141] Il termine di durata delle indagini preliminari è stato ampliato e l'eventuale proroga viene adottata dal giudice senza coinvolgimento della persona sottoposta alle indagini e/o della persona offesa (artt. 406, comma 5-*bis* e 407, comma 2, n. 4, c.p.p.); viene prevista una duplice presunzione di sussistenza delle esigenze cautelari e di adeguatezza della sola misura della custodia in carcere in caso di attivazione della procedura incidentale *de libertate* per l'applicazione di una misura cautelare nei confronti della persona sottoposta alle indagini (art. 275, comma 3, c.p.p.).

[142] Attualmente l'art. **3 del d.l. n. 374/2001**, a seguito degli interventi della l. 15 dicembre 2001, n. 438 e della l. 14 febbraio 2003, n. 34, reca la seguente formulazione *" Nei procedimenti per i delitti previsti dagli articoli 270*-ter ***e*** *280*-bis *del codice penale e per i delitti di cui all'articolo 407, comma 2, lettera a), n. 4 del codice di procedura penale, si applicano le disposizioni di cui all'articolo 13 del decreto-legge 13 maggio 1991, n. 152, convertito, con modificazioni, dalla legge 12 luglio 1991, n. 203*". Il n. 4 del comma 2, lett. a dell'art. 407 c.p.p. si riferisce ai delitti commessi per finalità di terrorismo o di eversione dell'ordinamento costituzionale per i quali la legge stabilisce la pena della reclusione non inferiore nel minimo a cinque anni o nel massimo a dieci anni, nonché delitti di cui agli articoli 270, terzo comma, e 306, secondo comma, del codice penale.

[143] Cfr. vigente art. 226 n. att. coord. trasns. c.p.p.

[144] SALVINI, *L'associazione finalizzata al terrorismo internazionale: problemi di definizione e prova della finalità terroristica*, in Cass. pen., 2006, p. 3366.

[145] Cass. Pen. Sez. VI, 30 gennaio 1996, *Bendebka*, in Giust. pen., 1997, c. 158; Id., 1° marzo 1996, *Ferdjani*, in Foro it., 1996, II, c. 578; Id., 1° giugno 1999, *Abdaoui Youssef*, in Dir. pen. proc., 2000, p. 485.

[146] SALVINI, *L'associazione finalizzata al terrorismo,* cit., p. 3367.

[147] ROSI, *Terrorismo internazionale: anticipazione della tutela penale e garanzie giurisdizionali*, in Dir. pen. proc., 2008, p. 458.

nell'associazione terroristica internazionale. In quest'ultima frequente è la distanza spaziale tra singolo o singola cellula all'inizio della catena operativa e l'obiettivo da colpire da altri; spesso quest'ultimo è ubicato in diverso luogo e ancora non deciso, predefinito e conoscibile in anticipo, in quanto la definizione del progetto delittuoso finale coinvolge numerosi militanti, organizzatori e ispiratori in diversi continenti, appartenenti anche a gruppi diversi[148]. Inoltre, nella realtà fenomenica del terrorismo internazionale, soprattutto di matrice islamica, di solito l'associazione ha una struttura cellulare che si forma e rimodula diversamente in differenti contesti spazio-territoriali, nei quali gli incontri fisici tra i partecipi del gruppo possono essere sporadici e realizzati sulla rete *internet* [149].

A seguito della novella in commento la giurisprudenza ha ricavato, per le associazioni con finalità di terrorismo internazionale, la natura di associazione terroristica non solo dall'inclusione dell'organizzazione negli elenchi di associazioni terroristiche stilati dagli organismi sovranazionali, ma anche dalla disamina del concreto manifestarsi dell'organizzazione alla stregua degli *indici descrittivi fattuali* fissati dall'art. 270-*sexies* c.p.[150]. Il delitto *ex* art 270-*bis* c.p. è configurabile in presenza di una struttura criminale che si prefigga la realizzazione di atti violenti qualificati dalla finalità di terrorismo anche internazionale provvista della capacità di dare agli stessi effettiva realizzazione, non essendo sufficiente una mera attività di proselitismo ed indottrinamento, finalizzata ad inculcare una visione positiva del martirio per la causa islamica e ad acquisire generica disponibilità ad unirsi ai combattenti in suo nome[151]. In presenza di una struttura organizzata sia pure in modo rudimentale il delitto di partecipazione ad un'associazione con finalità di terrorismo anche internazionale o di eversione dell'ordine democratico, di cui all'art. 270-*bis* cod. pen., è ritenuto configurabile con una condotta di adesione ideologica che si sostanzi in seri propositi criminali diretti alla realizzazione delle finalità associative, senza che sia necessario, data la natura di reato di pericolo presunto, che si abbia l'inizio di materiale esecuzione del programma criminale[152].

Una seconda fase di significative innovazioni normative rispetto all'assetto interno del diritto e della procedura penale per potenziare il contrasto del terrorismo è riferibile al **DL 27 luglio 2005, n. 144**, conv. con mod. nella *l*. 31 luglio 2005, n. 155.

Il testo normativo ha operato mutamenti di procedura penale di portata generale. Sono state introdotte nuove norme per le *intercettazioni preventive* cui all'art. 226 delle norme di att., coord. e trans. del c.p.p. (art. 4), sui *dati del traffico telefonico e telematici* (art. 6) sull'identificazione personale quanto al *fermo di identificazione* e ai *prelievi biologici coattivi* (art. 10) nonché in materia di *arresto e di fermo* (art. 13). La parte più rilevante degli interventi del 2005 è orientata al rafforzamento operativo degli apparati di polizia e di *intelligence:* possibilità dei *colloqui investigativi* con persone detenute o internate informazioni utili per lo svolgimento di indagini in materia di terrorismo; *permesso di soggiorno a fini investigativi*, da rilasciare agli stranieri che abbiano collaborato con l'autorità rendendo informazioni utili ai fini della prevenzione o dell'accertamento dei reati; introduzione di norme più rigorose di *regolamentazione di talune attività sottoposte ad autorizzazione amministrativa* (come gli esercizi pubblici di telefonia e *internet*; l'attività di volo; i servizi di vigilanza non richiedenti l'impiego di personale di forze di polizia, ecc.); nuove norme in materia di *espulsione degli stranieri per motivi di prevenzione del terrorismo*; attribuzione ai direttori dei servizi segreti della legittimazione a richiedere al procuratore generale presso la corte d'appello l'autorizzazione per lo svolgimento di *intercettazioni preventive*; obbligo di *identificazione degli acquirenti di*

[148] SALVINI, *L'associazione finalizzata al terrorismo*, cit., p. 3383.

[149] Nel senso che le strutture cellulari proprie delle associazioni criminose di matrice terroristica islamica sono caratterizzate da estrema flessibilità interna, così da rimodularsi secondo le pratiche esigenze che di volta in volta si presentano, oltre che in condizioni di operare anche contemporaneamente in più Stati, o in tempi diversi, con contatti fisici, telefonici o comunque a distanza anche sporadici Cfr. Cass Sez. fer., 18 agosto 2009, n. 34180, in Cass. pen., 2010, p. 3411.

[150] Così Cass. Pen. Sez. 5, n. 10380/2019, Rv. 277239 in fattispecie in tema di riconoscimento dell'"IS" come associazione terroristica operante in una dimensione spaziale globale, che si avvale di strutture dislocate in vari Paesi, protesa all'affermazione della "jihad globale" e finalizzata a commettere atti di violenza stragista per destabilizzare i pilastri degli ordinamenti costituzionali degli Stati e per attentare, in maniera indiscriminata e imprevedibile, alla vita ed integrità delle persone.

[151] Così Cass. Pen. Sez. 5, n. 48001/2016, Rv. 268164 in fattispecie in cui la Corte ha ritenuto insussistente il delitto di cui all'art. 270-*bis* c.p., individuando una serie di indici della limitata operatività del gruppo e sottolineando come l'attività di mero proselitismo e indottrinamento, potendo costituire precondizione ideologica per la costituzione di un'associazione terroristica, è valutabile ai fini dell'applicazione di misure di prevenzione

[152] Così Cass. Pen. Sez. 2, n. 24994/2006, Rv. 234345.

schede elettroniche per telefonia mobile, di *conservazione dei dati del traffico telefonico* e telematico e *nuovo regime di acquisizione* dei relativi dati a fini processuali.

Il legislatore, per limitare il rischio di contrasti interpretativi, ha declinato in termini normativi la **nozione di «*condotta con finalità di terrorismo*» (art. 270-*sexies* c.p.)**, introdotto le nuove fattispecie di *arruolamento* (artt. 270-*quater* [153]) e di *addestramento* con finalità di terrorismo anche internazionale (art. 270-*quinquies* c.p.[154]) ed esteso ai fatti di terrorismo internazionale istituti già previsti per il contrasto della

[153] Per Cass. pen. sez. 4, n. 23828/2019 l'art. *270-quater* c.p. è stata introdotta per assicurare un più efficace contrasto al fenomeno del terrorismo, soprattutto internazionale e di origine fondamentalista islamica, punendo la condotta di chi «al di fuori dei casi di cui all'articolo *270-bis, arruola* una o più persone per il compimento di atti di violenza ovvero di sabotaggio di servizi pubblici essenziali, con finalità di terrorismo, anche se rivolti contro uno Stato estero, un'istituzione o un organismo internazionale». Tale norma incriminatrice - al pari di quella contenuta nel successivo art. 270-*quinquies* c.p., che punisce *l'addestramento* ad attività con finalità di terrorismo anche internazionale - ha l'evidente scopo di ampliare lo spettro dell'intervento punitivo statuale, in quanto è destinata, per un verso, a colpire condotte con finalità di terrorismo anche internazionale poste in essere in Italia da soggetti che non risultino aver aderito ad una associazione *ex* art. 270-*bis* cod. pen., così finendo per garantire una forma di *anticipazione della tutela penale*; per altro verso, ha la finalità di evitare che autori di condotte di arruolamento che non espongono direttamente lo Stato italiano ad un pericolo di guerra, potessero rimanere impuniti, non essendo integrati gli estremi dei reati in materia di «arruolamento» previsti dagli artt. 244 e 288 c.p., né quelli del delitto di «reclutamento» di cui all'art. 4 della L. n. 210 del 1995. Per Cass. Pen., Sez. 1, n. 40699 del 09/09/2015, *Elezi*, Rv. 264719 la *nozione* di "*arruolamento*" con finalità di terrorismo anche internazionale è equiparabile a quella di "*ingaggio*", per esso intendendosi il raggiungimento di un '*serio accordo'* tra il soggetto che propone il compimento, in forma organizzata, di più atti di violenza ovvero di sabotaggio con finalità di terrorismo, e il soggetto che aderisce alla intesa. Ciò significa che, in mancanza del raggiungimento di un "accordo serio", ad esempio nel caso in cui vi sia stata una generica 'messa a disposizione' cui non sia seguita alcuna condotta indicativa di una concreta disponibilità ad entrare a far parte di una struttura gerarchica con le indicate finalità di terrorismo, sarebbe al più ipotizzabile l'applicazione di una misura di sicurezza a norma dell'art. 115 cod. pen. Per Cass. Pen. Sez. 2, n. 23618/2019, invece, per ritenere integrata la condotta di arruolamento passivo prevista dall'art. 270-*quater* comma 2 cod. pen. non è necessaria la prova del "serio accordo" con la associazione, ma è sufficiente la prova dell'*integrale disponibilità del neo-terrorista al compimento di tutte le azioni necessarie al raggiungimento degli scopi eversivi propagandati da Al Qaeda*; le condotte coperte dall'art. 270, comma 2, c.p. vanno identificate effettuando la diagnosi differenziale con la partecipazione all'associazione terroristica dovendosi verificato se l'individuo ha un preciso ruolo nell'organigramma dell'associazione terroristica, centrale o delocalizzata (nel qual caso vertendosi nell'ipotesi prevista dall'art. 270-*bis* c.p.), oppure abbia scelto di aderire al programma di Al Qaeda e di rendersi disponibile al compimento di atti connotati da finalità terroristiche, anche a progettazione individuale, ma comunque funzionali al raggiungimento degli obiettivi indicati dall'organizzazione (nel qual caso vertendosi nell'ipotesi prevista dall'art. 270-*quater*, comma 2, c.p.).

[154] Ai fini della configurabilità del reato di *addestramento ad attività con finalità di terrorismo anche internazionale*, commesso dalla persona che abbia acquisito autonomamente informazioni strumentali al compimento di atti con la suddetta finalità, è comunque necessario che il soggetto agente ponga in essere *comportamenti significativi sul piano materiale*, univocamente diretti alla commissione delle condotte di cui all'art. 270-*sexies* cod. pen., senza limitarsi ad una mera attività di raccolta di dati informativi o a manifestare le proprie scelte ideologiche (così Cass. Pen. Sez. 5 , n. 22066/2020, Rv. 279495 ha ritenuto configurabile in sede cautelare il reato di cui all'art. 270-*quinquies* c.p. sulla base di molteplici indici fattuali concreti, quali il possesso da parte dell'imputato di video ed immagini riconducibili alla propaganda terroristica per lo Stato islamico o illustrativi di tecniche per la preparazione di ordigni esplosivi, scaricati con elevata frequenza nell'arco di un significativo periodo di tempo, nonché di appunti manoscritti riproducenti la celebrazione di simboli e delle pratiche terroristiche dell'"Isis" e in cui l'indagato si proclamava "servo di Allah" votato al martirio; la partecipazione a chat di gruppo e canali di propaganda jihadista nei quali venivano manifestati propositi terroristici e di esaltazione del martirio e della guerra santa contro gli infedeli; il rinvenimento all'interno della sua abitazione di materiale destinato alla fabbricazione di un ordigno rudimentale; Cass. pen. Sez. 5, n. 6061/2017, Rv. 269581 ha ritenuto configurabile in sede cautelare il reato di cui all'art. 270-*quinquies* c.p. sulla base di molteplici indici fattuali concreti, tra i quali il possesso da parte dell'imputato di video ed immagini riconducibili alla propaganda terroristica per lo Stato islamico o illustrativi di tecniche per la preparazione di un ordigno, scaricati con elevata frequenza nell'arco di un significativo periodo di tempo, nonché l'avere in rubrica telefonica un'utenza collegata ad altra in uso a soggetto poi arrestato per detenzione di armi ed esplosivi; per contro Cass. Pen., Sez. 1, n. 7898/2019 ha ritenuto insufficiente ad integrare una condotta penalmente rilevante - ancorché dotata di valenza altamente sintomatica della contiguità con ambienti dell'estremismo islamico - l'avere l'imputato, tra l'altro, visionato numerosi video riguardanti tematiche "jihadiste", di cui alcuni strumentali all'auto-addestramento). L'art. 270-*quinquies* c.p. richiede un duplice dolo specifico, caratterizzato, non solo dalla realizzazione di una condotta in concreto idonea al compimento di atti di violenza ovvero di sabotaggio di servizi pubblici essenziali, ma anche dalla presenza di una delle finalità di terrorismo contemplate dall'art. 270-*sexies* cod. pen., le quali devono costituire oggetto di specifico accertamento sulla base delle emergenze del caso concreto (Cass. Pen. Sez. 1 , n. 7898/2020, Rv. 278499).

criminalità organizzata di stampo mafioso[155]. Di considerevole portata, in particolare, la **norma definitoria dell'art. 270-*sexies* c.p.** alla cui stregua «Sono considerate con finalità di terrorismo le condotte che, per la loro natura o contesto, **possono arrecare grave danno** ad un Paese o ad un'organizzazione internazionale e sono **compiute allo scopo di intimidire la popolazione o costringere** i poteri pubblici o un'organizzazione internazionale a compiere o astenersi dal compiere un qualsiasi atto **o destabilizzare o distruggere** le strutture politiche fondamentali, costituzionali, economiche e sociali di un Paese o di un'organizzazione internazionale, nonché le altre **condotte definite terroristiche o commesse con finalità di terrorismo da convenzioni** o altre norme di diritto internazionale vincolanti per l'Italia».

Come chiarito dalla giurisprudenza per integrare la finalità di terrorismo di cui all'art. 270-*sexies* c.p.., non basta che l'agente abbia intenzione di arrecare un grave danno al Paese ma occorre che la sua condotta crei la **possibilità concreta** - per la natura ed il contesto obiettivo dell'azione, nonché degli strumenti di aggressione in concreto utilizzati - che esso si verifichi, nei termini di un reale impatto intimidatorio sulla popolazione, tale da ripercuotersi sulle condizioni di vita e sulla sicurezza dell'intera collettività, posto che solo in presenza di tali condizioni lo Stato potrebbe sentirsi effettivamente coartato nelle sue decisioni[156]; non è sufficiente, dunque, la direzione dell'atteggiamento psicologico dell'agente, ma è necessario che la **condotta** posta in essere del medesimo sia **concretamente idonea a realizzare uno degli scopi** indicati nel predetto articolo (intimidire la popolazione, costringere i poteri pubblici a compiere o astenersi dal compiere un qualsiasi atto, destabilizzare o distruggere le strutture politiche fondamentali, costituzionali ecc. di un Paese o di un'organizzazione internazionale), determinando un evento di pericolo di portata tale da incidere sugli interessi dell'intero Paese[157].

La direttiva 2005/60/CE è stata recepita con il *D.lgs. 21 novembre 2007, n. 231,* per i profili concernenti il riciclaggio, e con il *D.lgs. 22 giugno 2007, n. 109* per la parte che attiene al finanziamento del terrorismo disciplina che completa, riordina e "*stabilizza*" la normativa preesistente, definendo in modo organico le strutture e le procedure con le quali le amministrazioni italiane partecipano alla lotta contro il finanziamento del terrorismo. In particolare, la nuova normativa ha ad oggetto l'introduzione delle «misure per prevenire l'uso del sistema finanziario a scopo di finanziamento del terrorismo e per attuare il congelamento dei fondi e delle risorse economiche per il contrasto del finanziamento del terrorismo e dell'attività di Paesi che minacciano la pace e la sicurezza internazionale in base alle risoluzioni delle Nazioni Unite o alle deliberazioni dell'Unione europea» (art. 2, comma 1, D.lgs. n. 109/2007) e la definizione della distribuzione delle competenze e l'assetto organizzativo interno predisposti per contrastare il finanziamento al terrorismo, distinguendo tra gli organi con compiti decisionali, che formulano le proposte di *listing* e *delisting*, e gli organi con compiti di vigilanza ed accertamento, ai quali è affidata l'attività istruttoria. In base all'originario art. 2, comma 4, del D.lgs. n. 231/2007 «ai fini del presente decreto per *finanziamento del terrorismo* vale la *definizione* di cui all'art. 1, comma 1, lettera a), del decreto legislativo 22 giugno 2007, n. 109»*;* quest'ultima esplicitava la definizione nei seguenti termini: «qualsiasi attività *diretta*, con qualsiasi mezzo, alla raccolta, alla provvista, all'intermediazione, al deposito, alla custodia o all'erogazione di fondi o di risorse economiche, in qualunque modo realizzati, destinati ad essere, in tutto o in parte, utilizzati al fine di compiere uno o più delitti con finalità di terrorismo o in ogni caso *diretti* a favorire il compimento di uno o più *delitti con finalità di terrorismo previsti dal codice penale*, e ciò indipendentemente dall'effettivo utilizzo dei fondi e delle risorse

---

[155] SALVINI, *I colloqui investigativi e i permessi di soggiorno a fini investigativi per il contrasto al terrorismo internazionale*, in Le nuove norme di contrasto al terrorismo, a cura di Dalia, Milano, 2006, p. 2.

[156] Cass. Pen. Sez. 1, n. 47479/2015, Rv. 265405; nella specie è stata esclusa la sussistenza della finalità di terrorismo negli episodi di danneggiamento ai cantieri TAV, ritenendo che le condotte delittuose non fossero concretamente idonee a costringere le pubbliche autorità a rinunciare alla realizzazione della linea ferroviaria ad alta velocità, né avessero la capacità di produrre un grave danno al Paese; il riferimento al "contesto", contenuto nel citato art. 270-*sexies*, e sulla base del quale deve essere valutato il significato della condotta, impone di dar rilievo al pericolo del "grave danno" anche quando questo non dipenda solo dall'azione individuale considerata, ma sia piuttosto il frutto dell'innesto di essa in una più ampia serie causale non necessariamente controllata dall'agente, fermo restando che questi deve rappresentarsi e volere tale interazione.

[157] Così Cass. Pen. Sez. 6, n. 28009/2014 Rv. 2600761; Id, n. 25949 /2008, Rv. 240465 per la quale non ricorre la circostanza aggravante della finalità di terrorismo prevista dall'art. 270-*sexies* c.p. nei fatti di devastazione commessi, in occasione della morte di un tifoso di calcio, da un gruppo di altri tifosi e concretatisi in aggressioni violente alle forze di polizia, lancio di bombe carta, assalto a caserme e incendio di autobus della stessa polizia, danneggiamento indiscriminato di auto e moto in sosta, in quanto in tali condotte, quantunque gravi, non è ravvisabile, in assenza di elementi di più adeguata strutturazione, la prospettiva teleologica ineludibile nella finalità medesima.

economiche per la commissione dei delitti anzidetti». Venendo operato un rinvio alle *norme penali*, a differenza di quanto avviene per il riciclaggio (per il quale è formulata un'autonoma definizione avente rilievo ai fini amministrativi), è necessario tener presente tutte le fattispecie rilevanti in materia di terrorismo, comprese quelle introdotte da recenti successivi interventi normativi. La portata delimitativa del rinvio, invero, è contraddetta dal fatto che tra i delitti con finalità di terrorismo l'art. 270-*sexies* c.p. annovera, con ampia estensione, anche «*le altre condotte definite terroristiche o commesse con finalità di terrorismo da convenzioni o altre norme di diritto internazionale vincolanti per l'Italia*» richiamando indirettamente i reati previsti dall'art. 2 della Convenzione di New York del 1999 che, a sua volta, al par. 1 lett. a) opera ulteriori rimandi anche alle disposizioni incriminatrici contenute nelle Convenzioni alla stessa allegate[158]. La definizione mira ad anticipare l'interesse verso operazioni che, *prescindendo dall'utilizzo effettivo dei fondi e delle risorse economiche* per la commissione di delitti connotati dalla finalità di terrorismo, risultano rivelatrici dell'orientamento finalistico ("*dirette*") «*alla raccolta, alla provvista, all'intermediazione, al deposito, alla custodia o all'erogazione*» delle stesse, per agevolare il compimento di uno o più delitti con finalità di terrorismo previsti dal codice penale; oltre a confermare l'irrilevanza dell'origine lecita o illecita dei fondi e delle risorse economiche è ribadito l'interesse verso il loro impiego per la realizzazione di un'attività illecita (*money dirtying*); lo palesano anche i concetti di "*fondi"* («le attività ed utilità finanziarie di qualsiasi natura, possedute anche per interposta persona fisica o giuridica», poi analiticamente esemplificati) e di "*risorse economiche*" («le attività di qualsiasi tipo, materiali o immateriali, mobili o immobili, ivi compresi gli accessori, le pertinenze e i frutti, che non sono fondi ma che possono essere utilizzate anche per interposta persona fisica o giuridica per ottenere fondi, beni o servizi»). Tali nozioni hanno rappresentato anche parametri di riferimento per l'adempimento degli obblighi di collaborazione passiva e attiva fissati dal D.lgs. n. 231/2007, secondo l'approccio basato sul rischio, collegato alla prevenzione della commissione delle diverse fattispecie penali contenute nel Libro II, titolo I, capitolo I e II del codice penale[159].

Con il "**pacchetto" antiterrorismo del 2015**, il legislatore italiano ha inteso attuare obblighi di diritto internazionale derivanti dalla Risoluzione n. 2178/2014 del Consiglio di Sicurezza delle Nazioni Unite[160]. La riforma ha inciso sulle norme in materia di terrorismo autorizzando un *generale arretramento della soglia di rilevanza penale, rendendo punibili alcuni atti meramente preparatori*. In dettaglio, dopo l'attacco terroristico alla redazione del giornale satirico *Charlie Hebdo* a Parigi, il

[158] Si tratta, in particolare delle seguenti Convenzioni: 1. Convenzione per la repressione dell'illecito sequestro di aeromobili (L'Aja, 16 dicembre 1970); 2. Convenzione per la repressione di atti illeciti diretti contro la sicurezza dell'aviazione civile (Montreal, 23 settembre 1971); 3. Convenzione sulla prevenzione e repressione dei reati contro le persone che godono di protezione internazionale, compresi gli agenti diplomatici (adottata dall'Assemblea Generale delle Nazioni Unite il 14 dicembre 1973); 4. Convenzione internazionale contro la cattura di ostaggi (adottata dall'Assemblea Generale delle Nazioni Unite il 17 dicembre 1979); 5. Convenzione internazionale sulla tutela del materiale nucleare (Vienna, 3 marzo 1980); 6. Protocollo per la repressione di atti illeciti di violenza negli aeroporti utilizzati dall'aviazione civile internazionale, complementare alla Convenzione per la repressione di atti illeciti diretti contro la sicurezza dell'aviazione civile (Montreal, 24 febbraio 1988); 7. Convenzione per la repressione di atti illeciti diretti contro la sicurezza della navigazione marittima (Roma, 10 marzo 1988); 8. Protocollo per la repressione di atti illeciti contro la sicurezza delle piattaforme fisse situate sulla piattaforma continentale (Roma, 10 marzo 1988); 9. Convenzione internazionale per la repressione di attentati terroristici perpetrati con esplosivo (adottata dall'Assemblea Generale delle Nazioni Unite il 15 dicembre 1997).

[159] Nella versione coeva all'introduzione del DL n. 109/2007 erano annoverabili tra essi l'associazione con finalità di terrorismo anche internazionale o di eversione dell'ordine democratico *ex* art. 270-*bis* c.p. (come riformulata dal DL n. 374/2001, convertito con modificazioni nella l. n. 438/2001), l'assistenza agli associati *ex* art. 270-*ter* c.p. (delitto introdotto dal DL n. 374/2001, cit.), l'arruolamento con finalità di terrorismo anche internazionale *ex* art. 270-*quater* c.p., l'addestramento ad attività con finalità di terrorismo anche internazionale *ex* art. 270-*quinquies* c.p. (fattispecie inserite dal DL n. 144/2005 convertito con modificazioni nella l. n. 155/2005), l'attentato per finalità terroristiche o di eversione *ex* art. 280 c.p., l'atto di terrorismo con ordigni micidiali o esplosivi *ex* art. 280-*bis* c.p. e il sequestro di persona a scopo di terrorismo o di eversione *ex* art. 289-*bis* c.p.

[160] La Risoluzione n. 2178 del 2014 del Consiglio di Sicurezza dell'Organizzazione delle Nazioni Unite, adottata ai sensi del Capo VII della Carta delle Nazioni Unite, ha evidenziato punti nevralgici affidandone l'attuazione agli Stati Membri: il contrasto all'estremismo violento; la cooperazione internazionale finalizzata ai controlli sul movimento dei sospetti terroristi, mediante l'eventuale utilizzo di misure di prevenzione; la risposta punitiva, consistente sostanzialmente nell'anticipazione della tutela penale, attraverso l'incriminazione degli stessi atti preparatori, antecedenti alla commissione di un attentato terroristico; l'espressa considerazione dei c.d. *foreign terrorist fighters*, prevedendo specifici obblighi di incriminazione del fenomeno in capo agli Stati Membri.

legislatore è intervenuto con il **DL n. 7 del 18 febbraio 2015**, convertito con modificazioni con L. n. 43 del 17 aprile 2015; dall'altro, mediante la L. del 28 luglio 2016, n. 153, sono state ratificate cinque Convenzioni internazionali in materia di prevenzione e contrasto al terrorismo [161], introducendo nuove ipotesi delittuose nel codice penale, che vanno ad aggiungere alle altre disposizioni codicistiche mirate alla lotta al terrorismo[162].

In particolare, in attuazione della Risoluzione 2178/2014, il DL n. 7/2015 ha introdotto ed esteso norme penali modificando la normativa in materia di terrorismo per incriminare le condotte dei *foreign terrorist fighters* ed ha ampliato i poteri di polizia nonché le situazioni nelle quali sono applicabili misure di prevenzione. In particolare, l'art. 1 della L. 43/2015 ha coniato nuove fattispecie di delitto in materia di terrorismo, aggiungendo un secondo comma all'art. 270-*quater* c.p.[163], introducendo l'art. 270-*quater*.1 c.p.[164], modificando l'art. 270-*quinquies* c.p.[165] e, da ultimo, aggiungendo la pena accessoria della perdita della potestà genitoriale, qualora sia coinvolto un minore, nel caso di condanna per i delitti *ex* artt. 270-*bis*, 270-*ter*, 270-*quater*, 270-*quater*.1 e 270-*quinquies* c.p.

La *l. n. 153 del 28 luglio 2016*, contestualmente alla ratifica delle ricordate Convenzioni internazionali, ha operato ulteriori innovazioni nel Codice penale, introducendo o ulteriori fattispecie di reato nell'ambito delle norme volte al contrasto del terrorismo[166].

In primo luogo, ha inserito l'art. 270-*quinquies*.1 c.p., che punisce chiunque, al di fuori dei casi di cui agli articoli 270-*bis* e 270-*quater*.1, raccoglie, eroga o mette a disposizione beni o denaro, in qualunque modo realizzati, destinati a essere in tutto o in parte utilizzati per il compimento delle condotte con finalità di terrorismo di cui all'art. 270-*sexies*, indipendentemente dall'effettivo utilizzo dei fondi per la commissione delle citate condotte, con la reclusione da sette a quindici anni. Il secondo comma della disposizione prevede, inoltre, che chiunque deposita o custodisce i beni o il denaro indicati al primo comma è punito con la reclusione da cinque a dieci anni[167]. In secondo luogo, vengono introdotti gli artt. 270-*quinquies*.2 e 270-*septies* c.p.,

---

[161] L. 28 luglio 2016, n. 153 intitolata "Norme per il contrasto al terrorismo, nonché ratifica ed esecuzione: a) della Convenzione del Consiglio d'Europa per la prevenzione del terrorismo, fatta a Varsavia il 16 maggio 2005; b) della Convenzione internazionale per la soppressione di atti di terrorismo nucleare, fatta a New York il 14 settembre 2005; c) del Protocollo di Emendamento alla Convenzione europea per la repressione del terrorismo, fatto a Strasburgo il 15 maggio 2003; d) della Convenzione del Consiglio d'Europa sul riciclaggio, la ricerca, il sequestro e la confisca dei proventi di reato e sul finanziamento del terrorismo, fatta a Varsavia il 16 maggio 2005; e) del Protocollo addizionale alla Convenzione del Consiglio d'Europa per la prevenzione del terrorismo, fatto a Riga il 22 ottobre 2015".

[162] Per un'analisi critica delle soluzioni individuate nel tempo dal legislatore italiano e sulla loro compatibilità con i principi costituzionali di legalità, offensività, materialità e colpevolezza cfr. REITANO, *Riflessioni in margine alle nuove fattispecie antiterrorismo*, Rivista Italiana di Diritto e Procedura Penale 2007, pp. 217ss.; MARINO, *Il sistema antiterrorismo alla luce della l. 43/2015: un esempio di "diritto penale del nemico"*? ibid., fasc. 3, 2016, p. 1388.

[163] Tale norma punisce con la pena della reclusione da cinque a otto anni , fuori dai casi di cui all'articolo 270-*bis* c.p. e salvo il caso di addestramento (art. 270-*quinquies* c.p.), la persona che viene arruolata dai soggetti che commettono il delitto di cui al primo comma; in tal modo sono incriminati i c.d. "*lupi solitari*" , soggetti i quali, pur non facendo tecnicamente parte delle associazioni di cui all'art. 270-*bis* c.p., arruolano e vengono arruolati al fine di commettere atti terroristici. In tema di arruolamento con finalità di terrorismo anche internazionale, la nozione di "arruolamento" è stata ritenuta equiparabile a quella di "ingaggio", per esso intendendosi il raggiungimento di un serio accordo tra soggetto che propone il compimento, in forma organizzata, di più atti di violenza ovvero di sabotaggio con finalità di terrorismo e soggetto che aderisce (Cass. Pen. Sez. 1, n. 40699/2015, Rv 264719-01).

[164] Si tratta della fattispecie di "organizzazione di trasferimenti per finalità di terrorismo" che punisce, fuori dai casi di cui agli artt. 270-*bis* e 270-*quater*, chiunque *organizza*, *finanzia* o *propaganda* viaggi in territorio estero, finalizzati al compimento delle condotte di cui all'art. 270-*sexies* c.p., con la pena della reclusione da cinque a otto anni. Sulla problematicità della figura criminosa e sul suo "risicato" ambito di applicazione cfr. MARINO, o*p. cit*., pp. 1420 ss.

[165] Il terzo comma dell'art. 1 della l. n. 43 del 2015 modifica l'art. 270-*quinquies* c.p. ampliandone l'ambito di applicazione: accanto a colui il quale addestra o comunque fornisce istruzioni sulla preparazione o sull'uso di materiali esplosivi, di armi da fuoco o di altre armi, di sostanze chimiche o batteriologiche nocive o pericolose, nonché di ogni altra tecnica o metodo per il compimento di atti di violenza ovvero di sabotaggio di servizi pubblici essenziali, con finalità di terrorismo, ed alla persona addestrata, la norma punisce, adesso, anche la persona che avendo acquisito, anche autonomamente, le istruzioni per il compimento degli atti di cui al primo periodo, pone in essere comportamenti *univocamente* finalizzati alla commissione delle condotte di cui all'art. 270-*sexies* c.p.

[166] Per una ricostruzione critica, cfr. MARINO, *op. cit*., pp. 1389-1426.

[167] La norma, finalizzata a punire il finanziamento di condotte con finalità di terrorismo, dovrebbe colmare una lacuna del sistema, incriminando i finanziatori di atti terroristici organizzati al di fuori di un'associazione *ex* art. 270-*bis* c.p.

disposizioni collegate dal punto di vista finalistico, che mirano a reprimere i focolai terroristici alla base, sottraendo loro le disponibilità materiali ed economiche. Viene, perciò, incriminato chiunque ponga in essere condotte di sottrazione di beni o denaro sottoposti a sequestro per finalità di prevenzione del finanziamento delle condotte terroristiche; inoltre è prevista l'obbligatorietà, alle condizioni dell'art. 270-*septies* c.p., della confisca dei beni, utilizzati o destinati al compimento di reati di terrorismo, nel caso di condanna per taluno dei delitti con la finalità di cui all'art. 270-*sexies* c.p. Infine, la riforma ha inserito l'art. 280-*ter* c.p., rubricato "*atti di terrorismo nucleare",* disposizione che contiene in realtà due diverse incriminazioni, correlate strettamente alla finalità di cui all'art. 270-*sexies*: è punito con la pena della reclusione non inferiore a quindici anni chiunque procura a sé o ad altri materia radioattiva, ovvero crea un ordigno nucleare o ne viene altrimenti in possesso; e con la reclusione non inferiore a vent'anni, chiunque utilizza materia radioattiva o un ordigno nucleare ovvero utilizza o danneggia un impianto nucleare in modo tale da rilasciare o con il concreto pericolo che rilasci materia radioattiva. Il terzo comma dell'art. prevede, infine, che le disposizioni di cui ai primi commi si applicano altresì quando la condotta ivi descritta abbia ad oggetto materiali o aggressivi chimici o batteriologici.

La forte anticipazione della tutela e la soggettivizzazione delle fattispecie, accomunate dalla direzione finalistica della condotta, intensificano il pericolo di rarefazione dell'offensività del fatto tipico[168]; in particolare, nelle nuove fattispecie di arruolamento, addestramento, istruzione, organizzazione di viaggi, il dolo specifico, che dovrebbe qualificare il disvalore del fatto, fatica ad assolvere la funzione penalmente tipizzante, perché sono incriminate condotte molto anticipate e distanti da quelle terminali provviste di effettiva dimensione lesiva[169].

Di rilievo, altresì, le modifiche in tema di destinatari delle misure di prevenzione personali e patrimoniali per persone di interesse intervenute tra il 2015 e il 2017 estesa la categoria «agli *indiziati* di uno dei reati previsti dall'art. 51, comma 3-*quater*, del codice di procedura penale e a *coloro che, operanti in gruppi o isolatamente, pongano in essere atti preparatori, obiettivamente rilevanti*, ovvero esecutivi diretti a sovvertire l'ordinamento dello Stato, con la commissione di uno dei reati previsti dal capo I del titolo VI del libro II del codice penale o dagli articoli 284, 285, 286, 306, 438, 439, 605 e 630 dello stesso codice, nonché alla commissione dei reati con finalità di terrorismo anche internazionale ovvero a prendere parte ad un conflitto in territorio estero a sostegno di un'organizzazione che persegue le finalità terroristiche di cui all'art. 270-*sexies* del codice penale»[170].

- 6.5.2. La lotta al terrorismo internazionale, tra disciplina di contrasto della criminalità organizzata e del riciclaggio: dal fenomeno finanziario a quello ideologico.

Le tecnologie di IA sono sempre più apprezzate per le opportunità di impiego nella repressione del terrorismo, in particolare per le loro peculiari capacità di raccolta, conservazione e analisi di dati

---

[168] Di «*offensività debole*» parla MILITELLO, *La prevenzione del terrorismo*, in Dir. pen. cont. - Riv. trim., 2017, p. 6; nella stessa prospettiva l'analisi di Palazzo, *Nemico-nemici-nemico: una sequenza inquietante per il futuro del diritto penale*, in RIDPP 2/2020, p. 708, il quale sottolinea come in materia di terrorismo internazionale, il legislatore italiano ma prima ancora quello europeo siano addivenuti ormai alla formulazione di fattispecie monosoggettive e marcatamente preparatorie radicalmente prive di capacità destabilizzante delle istituzioni statali, carenti di reale offensività, in quanto pressoché interamente imperniate sulla relazione nemicale con l'autore che affonda le sue radici quasi nella costituzione antropologica del terrorista. Osserva l'A.: «l'ostilità dell'ordinamento si giustifica essenzialmente sull'attributo di disumanità che ne caratterizza la personalità: la sua scelta di mirare a vittime innocenti, inermi ed indiscriminate, al di là di quale sia la effettiva, oggettiva pericolosità dell'atto compiuto, è quanto basta per costituirlo — ideologicamente — come nemico dell'ordine assiologico istituzionale».

[169] CAVALIERE, *Considerazioni critiche intorno al d.l. antiterrorismo, n. 7 del 18 febbraio 2015*, in Dir. pen. cont., 31 marzo 2015; LEO, *Nuove norme in materia di terrorismo*, in Dir. pen. cont., 19 ottobre 2016, pp. 5 ss.; PECCIOLI, *Punibilità di atti preparatori alla realizzazione di condotte terroristiche*, in Studium iuris, 2015, pp. 770 ss.

[170] La lettera d) dell'art. 4 del d.lgs. n. 159/2011 è stata modificata dall'art. 4, comma 1, lett. a), DL 18 febbraio 2015, n. 7, convertito, con mod., dalla l. 17 aprile 2015, n. 43 e poi sostituita dall'art. 1, comma 1, lett. b), l. 17 ottobre 2017, n. 161; per osservazioni critiche sull'ulteriore anticipazione dell'intervento prevenzionale, cfr. PELISSERO, *La legislazione antiterrorismo. il prototipo del diritto penale del nemico tra garanzie e rischi di espansione* in Rivista Italiana di Diritto e Procedura Penale, fasc. 2, 1/6/2020, p. 745; BARTOLI, *Legislazione e prassi in tema di contrasto al terrorismo internazionale: un nuovo paradigma emergenziale?* in Dir. pen. cont. - Riv. trim., 2017, p. 236.

utili al monitoraggio delle attività sospette.

La relazione tra lotta al terrorismo e contrasto alla criminalità organizzata scaturisce dalla stretta connessione dei due fenomeni[171], che ricorre anche rispetto al riciclaggio. Ciò importa che le tecniche di contrasto dei tre fenomeni, almeno per la **componente finanziaria**, registrano significativi momenti di comunanza[172].

I due "**pilastri**" della normativa di prevenzione e contrasto al finanziamento del terrorismo sono in effetti rappresentati da: *i*) **misure di «congelamento»**, consistenti nel blocco operativo di capitali e risorse economiche posseduti o controllati anche indirettamente da terroristi designati in apposite liste, dirette ad interrompere i flussi destinati al finanziamento del terrorismo, agendo dunque sulle fonti; *ii)* **misure analoghe a quelle antiriciclaggio** (registrazione delle operazioni, adeguata verifica, segnalazioni operazioni sospette, organizzazione e controlli) volte a ostacolare l'utilizzo del sistema finanziario per il finanziamento del terrorismo.

Sul piano finanziario gli strumenti e le tecniche finora adottati vanno mantenuti e perfezionati perché continuino a essere un valido presidio contro minacce analoghe a quelle passate; nondimeno, devono essere affiancati da approcci nuovi per contrastare le diverse tipologie di minaccia e accrescere la capacità delle informazioni finanziare di individuare rischi di terrorismo. Sotto questo aspetto la **collaborazione** fra autorità rappresenta un nodo essenziale, perché il contrasto finanziario deve organicamente inserirsi e integrarsi in una politica di prevenzione e contrasto multidisciplinare: militare, politico/diplomatica, di *law enforcement* e *intelligence*), di prevenzione sociale (dialogo interculturale e interreligioso). In tal senso l'UIF ha incremento professionalità «tecniche» per il trattamento di Big Data, creando magazzini di dati integrati (*data warehouse*) e impiegato strumenti per "analisi di rete" (*network analysis*) maggiormente proattive, non più concentrate sulle sole SOS ma arricchite da informazioni e dati delle FIU estere, delle autorità giudiziaria, elementi informativi complementari (come ispezioni, anagrafe dei rapporti). Di interesse alcune esperienze di monitoraggio dei flussi finanziari verso paesi arabi e medio orientali, basati su dati delle segnalazioni antiriciclaggio aggregate e su informazioni richieste appositamente agli intermediari. Gli andamenti anomali identificati sono stati sottoposti ad approfondimento finanziario, anche interagendo con le FIU straniere competenti, per valutare il successivo inoltro alle autorità di competenza. Le linee di evoluzioni attese, passano dallo sviluppo di strumenti avanzati, Network analysis, classificazione immediata delle informazioni sulla base di tecniche semantiche e di estrazione di testi (*text mining*), attenzione all'utilizzo di valute virtuali, proattività nell'acquisizione di dati rilevanti, accertamenti mirati con la collaborazione degli intermediari, individuazione di pattern comportamentali**.** Sotto

[171] La Posizione comune 2001/931/PESC del Consiglio, del 27 dicembre 2001, relativa all'applicazione di misure specifiche per la lotta al terrorismo poneva in evidenza i profondi legami tra terrorismo e criminalità organizzata di tipo mafioso, come rilevato nel campo del contrabbando di merci e dei traffici di materiali da armamento. Del resto, negli anni '80, in Italia, le mafie hanno affermato la loro egemonia anche per mezzo di azioni terroristiche: alle stragi del 1993 sono seguite condanne per delitti ritenuti aggravati dall'aver agito per finalità di terrorismo ed eversione dell'ordine costituzionale (art. 1 d.l n. 625/1979, conv. dalla l. 15/1980); in tema F. Roberti, *Terrorismo internazionale. Contrasto giudiziario e prassi operative*; in *Gli Speciali di Questione Giustizia: Terrorismo internazionale. Politiche della sicurezza. Diritti fondamentali;* settembre 2016.

[172] Regolamento (UE) 2023/2131 del Parlamento Europeo e del Consiglio del 4 ottobre 2023 che modifica il regolamento (UE) 2018/1727 del Parlamento europeo e del Consiglio e la decisione 2005/671/GAI del Consiglio, per quanto riguarda lo **scambio digitale di informazioni nei casi di terrorismo**, **considerando 2**4: **«**Il terrorismo, la criminalità organizzata e le forme gravi di criminalità sono oggi fenomeni molto dinamici e globalizzati che spesso interessano due o più Stati membri. Sebbene il terrorismo avesse già una forte componente transnazionale, con l'uso e la disponibilità dei mezzi di comunicazione elettronica la collaborazione transnazionale tra terroristi è notevolmente aumentata. Il carattere transnazionale di un reato di terrorismo potrebbe non essere noto nel momento in cui il caso è deferito a un'autorità giudiziaria ma potrebbe emergere da un controllo incrociato di dati da parte di Eurojust. L'indagine o l'azione penale relativa a reati di terrorismo richiede pertanto il coordinamento e la cooperazione tra le autorità responsabili dell'azione penale o un'azione penale su basi comuni, come previsto dall'articolo 85 del trattato sul funzionamento dell'Unione europea (TFUE). È opportuno scambiare con Eurojust le informazioni sui casi di terrorismo in modo tempestivo a meno che le circostanze specifiche del caso non indichino chiaramente che hanno un carattere puramente nazionale».

questo profilo riveste un ruolo determinante l'esperienza degli analisti che trova formalizzazione negli indicatoridi anomalia, cui potrebbero essere applicate tecniche algoritmiche sul modello di quelle impiegati in alcuni progetti di ML già illustrati in precedente (cfr. Provvedimento recante gli indicatori di anomalia del Direttore dell'Unità di Informazione Finanziaria per l'Italia del 12.5.2023, Sezione C, Indicatore di anomalia n. 33[173])

- 6.5.3. Le nuove fenomenologie di terrorismo internazionale: impiego di applicazioni IA.

Dopo gli attentati dell'11 settembre 2001, le democrazie occidentali hanno dovuto fronteggiare una minaccia nuova, non assimilabile ai fenomeni terroristici conosciuti sino ad allora nei diversi Paesi europei; la nuova **fenomenologia terroristica internazionale** si è caratterizzata per **imprevedibilità e continua variazione** delle strategie impiegate per gli attentati, rendendo estremamente difficile prevenirli[174]. Ciò ha accresciuto l'importanza della **tecnologia** nel contrasto della minaccia terroristica internazionale, anche perché all'avanzamento della prima è corrisposto il preoccupante sviluppo della seconda.

Del resto, l'aumento dell'**attività terroristica "on line"** è una sfida crescente, tanto da divenire connotato tipico del terrorismo moderno. Come ricorda il rapporto congiunto sulla lotta al terrorismo on line con l'intelligenza artificiale, prodotto nel 2021 della *partnership* tra lo *United Nations Counter-Terrorism Centre presso lo United Nations Office of Counter-Terrorism* (UNOCT) e lo *United Nations Interregional Crime and Justice Research Institute* (UNICRI) attraverso il suo *Centre for Artificial Intelligence and Robotics*[175], nel 2020 Europol e 17 Stati Membri hanno identificato e valutato per la rimozione ben 1.906 URL[176] che collegavano a contenuti terroristici su 180 piattaforme e siti *web*. Facebook ha segalato di aver rimosso nel corso di due anni più di 26 milioni di contenuti di gruppi come lo Stato Islamico dell'Iraq e del Levante (ISIL) e Al-Qaeda. Negli ultimi sei mesi del 2018, ad esempio, YouTube ha rimosso circa 150.000 video contenenti riferimenti al terrorismo e il 98% di essi era stato segnalato dagli algoritmi a disposizione della piattaforma[177].

Internet e i *social media*, dunque, si profilano quali potenti strumenti nelle mani di questi gruppi, permettendo loro di comunicare, diffondere i loro messaggi, raccogliere fondi, reclutare sostenitori, ispirare e coordinare attacchi e prendere di mira persone vulnerabili. Il *report* congiunto UNOCT e UNICRI ha esaminato come l'IA possa essere utilizzata per combattere la minaccia del terrorismo *online*; riconoscendo la minaccia del terrorismo, i crescenti tassi di digitalizzazione e la fiorente popolazione giovane, vulnerabile e *online* nell'Asia meridionale e nel Sud-Est asiatico, il *report* fornisce una guida alle forze dell'ordine e alle agenzie antiterrorismo dell'Asia meridionale e del Sud-Est asiatico sulla potenziale applicazione dell'IA per contrastare il terrorismo *online*, così come sulle molte sfide tecniche, politiche e relative ai diritti umani che dovranno considerare e affrontare se dovessero scegliere di farlo. Nella Strategia Globale Antiterrorismo delle Nazioni Unite (A/RES/60/288), gli Stati Membri hanno deciso di lavorare con le Nazioni Unite con il dovuto riguardo alla **riservatezza**, rispettando i diritti umani e in conformità con altri **obblighi di diritto internazionale**, per esplorare modi per coordinare gli sforzi a livello internazionale e regionale per

[173]https://uif.bancaditalia.it/normativa/norm-indicatori-anomalia

[174] A. VEDASCHI, *Intelligenza artificiale e misure antiterrorismo alla prova del diritto costituzionale*; in *Consulta Online*, 17 febbraio 2020. Il terrorismo internazionale fenomeno dinamico e multiforme, che nel tempo ha subito diverse trasformazioni, nella struttura e nelle ambizioni, sino alla sfociata autoproclamazione dello Stato Islamico nel 2014. Per una ricostruzione delle vicende dello Stato islamico nel "*post*-Califfato", F. MARRONE-M. OLIMPIO, *La minaccia jihadista dopo il Califfato*, nelle pubblicazioni dell'ISPI (28 marzo 2019); A. VEDASCHI, *Da al-Qā'ida all'IS: il terrorismo internazionale* si è fatto *Stato?* in *Riv. trim. dir. pubbl.*, 1/2016, 41 ss.

[175] Reperibile su *https://unicri.it/News/-Countering-Terrorism-Online-with-Artificial-Intelligence.*

[176] *Uniform Resource Locator*, noto con l'acronimo URL (lett. "localizzatore uniforme di risorse"), è una sequenza di caratteri che identifica univocamente l'indirizzo di una risorsa su una rete di computer, come un documento, un'immagine, un video, tipicamente presente su un *host server* e resa accessibile a un *client.*

[177] S. WOJCICKI, *Expanding our work against abuse of our platform*, YouTube Official Blog, 4 December 2017.

contrastare il terrorismo in tutte le sue forme e manifestazioni su Internet e utilizzare Internet come strumento per contrastare la diffusione del terrorismo. Allo stesso tempo, la Strategia riconosce che gli Stati Membri possono richiedere assistenza per poter rispettare questi impegni.

Volendo esaminare le **prospettive di impiego dell'IA da parte del terrorismo**, gli approfondimenti più aggiornati a livello internazionale[178] propongono una tripartizione di macroaree: *i*) minacce informatiche (*cyber threats)*; *ii*) minacce fisiche (*psysical threats)*; *iii*) minacce politiche (*political threats*)[179]. In termini generali, il contrasto a tali fenomeni è governato dai legislatori preferendo all'approccio *repressivo* (volto a punire chi abbia già commesso il fatto di reato) quello **preventivo**, mirato ad una "tutela anticipatoria", finalizzata ad evitare che situazioni "pericolose" evolvano nell'effettiva commissione di atti criminali; l'azione a forte connotazione preventiva rappresenta la "**chiave di lettura"** delle attuali strategie di contrasto al terrorismo internazionale[180].

Nella **prima macroarea**, l'impiego dell'IA preoccupa per la capacità di moltiplicare i risultati di offesa, soprattutto rispetto ad aggressioni ordinarie quali quelle **DoS o DDos**[181]. Come si anticipava il contrasto di tali applicazioni passa dalla prevenzione e dalla gestione della sicurezza prima che dall'investigazione giudiziaria; le infrastrutture informatiche strategiche (siano quelle economiche o militari, sia il sistema dei trasporti, della salute e dell'istruzione) devono essere presidiate da livelli di protezione adeguati al grado della minaccia. Parimenti è a dirsi, per le altre tipologie di aggressioni informatiche (*malware*, *ransomware*, *man-in-the-middle*[182], ecc.), potenzialmente amplificabili dagli

---

[178] Cfr. il Rapporto "*Algorithms and terrorism: the malicious use of artificial intelligence for terrorist purposes*", redatto dal *Counter Terrorism* delle Nazioni Unite (UNCCT) e dal corrispondente centro di ricerca (UNICRI) nel 2021. Il documento (reperibile su *https://unicri.it/News/Algorithms-Terrorism-Malicious-Use-Artificial-Intelligence-Terrorist-Purposes*) mira a contribuire a comprendere il rischio che l'IA cada nelle mani dei terroristi, specie allorché le nuove tecnologie diverranno di accesso comune. Il rapporto cerca di comprendere in quale misura l'IA possa divenire un'alternativa a disposizione del terrorismo e quali conseguenze potrebbero derivarne per la comunità internazionale. Dopo aver fornito una panoramica generale, fornendo statistiche a conforto delle crescenti preoccupazioni tra gli esperti riguardo all'uso dannoso di questa tecnologia, anche da parte di terroristi (cap. I) ed offerto un quadro generale dell'IA (cap. II) , viene illustrata la potenziale minaccia rappresentata dall'utilizzo di tali nuove tecnologie da parte di gruppi e individui terroristici presentando diverse esemplificazioni nelle quali le tecnologie come *Internet* e i *social media* sono state strumenti potenti (cap. III). Il quarto capitolo cerca di contestualizzare ulteriormente l'uso dannoso dell'IA esaminando le tre categorie di minacce (*cyber*, fisiche e politiche) identificate nella letteratura. Il cap. V approfondisce la questione se il terrorismo reso possibile dall'IA possa essere una prospettiva realistica, il cap. VI fornisce una panoramica degli usi dannosi presenti e futuri dell'IA da parte di gruppi e individui terroristici, il cap. VII presenta tre scenari dell'utilizzo dell'IA per scopi terroristici, concentrandosi sull'individuazione delle password, nonché su *ransomware,* droni, riconoscimento facciale, *deepfake* e passaporti modificati. Il cap. VIII valuta se vi sia motivo di preoccupazione riguardo ai gruppi terroristici e agli individui che utilizzano direttamente l'IA per migliorare o amplificare un attacco. Il cap. IX conclude il rapporto offrendo una serie di raccomandazioni e suggerimenti per le diverse azioni di *follow-up* per rafforzare la capacità per prepararsi ad un futuro terrorismo basato sull'IA.

[179] La tripartizione è ripresa anche da M. ROMANELLI, *Intelligenza artificiale, influenza sul mercato politico e reati contro la personalità dello stato. La criminalità terroristica*, in Quaderno II/2022 della *Rivista trimestrale della Scuola di perfezionamento per le forze di polizia*, curato dal Direttivo della Fondazione Vittorio Occorsio, pubblicato anche su Sistema Penale, 2022.

[180] Per uno studio sul carattere preventivo delle diverse misure antiterrorismo, si veda C.C. MURPHY, *EU Counter-Terrorism Law: Pre-Emption and the Rule of Law*, Hart Publishing, Oxford, 2012

[181] L'acronimo sta per **Distributed Denial of Service**, traducibile in italiano come Interruzione distribuita del servizio, e consiste nel tempestare di richieste un sito, fino a metterlo ko e renderlo irraggiungibile. È tra gli attacchi che colpiscono un'impresa ogni cinque minuti insieme ai *malware* e ai *ransomware*. Il DoS, cioè Denial of Service è un'azione il cui obiettivo è ingolfare le risorse di un sistema informatico che fornisce un determinato servizio ai computer connessi. Ci riesce prendendo di mira server, reti di distribuzione, o data center che vengono inondati di false richieste di accesso, a cui non riescono a far fronte. I DDos funzionano allo stesso modo, ma avvengono su scala molto più ampia. Nel caso dei Dos, infatti, bisogna difendersi da una sola sorgente di traffico informatico: per esempio, un numero elevato di e-mail in arrivo contemporaneamente. Mentre **durante gli attacchi DDoS le domande fasulle arrivano nello stesso momento da più fonti**. Tutto ciò determina una maggiore efficacia dello strumento che per funzionare ha bisogno di minor tempo. Gli effetti disastrosi, invece, durano più a lungo: da qualche ora fino a diversi giorni, in base alla prontezza con cui si reagisce.

[182] In crittografia e nella sicurezza informatica un **attacco man in the middle** (MITM, MIM, MIM *attack* o MITMA, in italiano *"uomo nel mezzo"*) indica un attacco informatico in cui un intruso segretamente ritrasmette o altera la comunicazione tra due parti che ritengono di comunicare direttamente tra di loro. Con le intercettazioni (***eavesdropping)***

algoritmi di ML. Occorre considerare che l'aggressione informatica può realizzarsi con la distruzione (totale/parziale, duratura/limitata nel tempo) di un sistema informatico, ma anche nell'esfiltrazione di dati rilevanti, con successivi impieghi criminali dei dati.

Venendo alle ***Psysical threats***, l'IA può moltiplicare e potenziare gli attacchi terroristici; si pensi all'impiego di **droni o di autovetture a guida autonoma**. Uno dei vantaggi dell'uso di *self-driving car* è quello di risparmiare la vita del responsabile dell'attacco, o, in alternativa, di evitarne la cattura; ciò al netto di meccanismi di radicalizzazione che riconnettono al sacrificio individuale un valore, fortemente rivendicato e propagandato, come nel *jihadismo* globale. Uno spettacolare attacco attraverso la più avanzata tecnologia può diffondere la paura, il mezzo peculiare dell'agire terroristico.

Quanto alle ***Political threats*** il rischio di condizionamento della politica attraverso ***deepfake***[183] è molto elevato, con esempi significativi in numerose parti del mondo, restandone condizionate elezioni attraverso la distruzione dell'immagine pubblica di un candidato o di un partito. Inoltre, tale tecnologia può moltiplicare l'efficacia di discorsi d'odio, e quindi anche della **radicalizzazione** attraverso la rete, con l'ulteriore vantaggio di anonimizzazioni sicure. Il contenuto audio-video falso o manipolato, realizzato attraverso sistemi di IA, può avere efficacia devastante, per la qualità del falso, per la straordinaria capacità di disseminazione e per la velocità della stessa, tanto che nessuna correzione né alcun tentativo di eliminazione potrebbe realmente e completamente sanare l'effetto del falso. La propaganda *online* ha rappresentato uno dei più importanti fattori di successo del terrorismo c.d. islamico, e la rete rappresenta il "luogo" d'eccellenza per l'efficace funzionamento di meccanismi di radicalizzazione, cui l'evoluzione tecnologica ha ovviamente contribuito in modo decisivo.

- 6.5.4. Il contrasto dell'IA al terrorismo.

Venendo alle applicazioni dell'IA nel contrasto dei fenomeni terroristici, può essere premesso che la rapidissima evoluzione della tecnologia[184] registrata nei primi venti anni di questo secolo ha traslato i nuovi strumenti tecnologici digitali da una ristretta *élite* alla dimensione "pertinenziale" di ciascun individuo. Tale pervasività dell'utilizzo del mezzo tecnologico lo ha trasformato in una di "arma a doppio taglio": da una parte, è noto come i terroristi islamisti la sfruttino quale efficace mezzo di radicalizzazione, indottrinamento e reclutamento[185]; dall'altra, la tecnologia offre vantaggi alle autorità pubbliche impegnate nella lotta al terrorismo internazionale[186]. Il recente Regolamento

---

l'attaccante crea connessioni indipendenti con le vittime e ritrasmette i messaggi del mittente facendo credere loro di comunicare tramite una connessione privata, mentre l'intera conversazione è controllata dall'intruso, in grado di intercettare i i messaggi importanti e/o iniettarne di nuovi. In molte circostanze questo può avvenire all'interno di un WI-FI *access point* non criptato e l'attaccante può inserirsi come *"uomo nel mezzo"*. Con lo ***spoofing*** viene variamente impiegata la falsificazione dell'identità (*spoof*). L'attacco può funzionare solo se nessuna delle due parti è in grado di sapere che il collegamento che le unisce è stato effettivamente compromesso da una terza parte. La maggior parte dei protocolli di crittografia includono una qualche forma di autenticazione *endpoint* per prevenire attacchi MITM.

[183] È la tecnologia **che sfrutta le reti neurali e l'intelligenza artificiale per modificare filmati, audio o immagini in modo più o meno automatico**, ottenendo così nuovi contenuti "fasulli", ma estremamente realistici.

[184] Sull'impatto della tecnologia su un amplissimo novero di attività umane; cfr. P. COSTANZO, *Il fattore tecnologico e le sue conseguenze*, in *Rass. parl.*, 4/2012, 811 ss.

[185] La rete è usata anche per lo scambio di informazioni tra i membri delle "cellule del terrore" e per l'organizzazione di attentati; tali comunicazioni si svolgono solitamente sul c.d. *dark web*, una parte della rete che non viene indicizzata dagli ordinari motori di ricerca e che necessita di chiavi di accesso particolari. Sul punto v. G. WEIMANN, *Going Dark: Terrorism on the Dark Web*, in 39 *Studies in Conflicts and Terrorism*, 2016, 195 ss.

[186] La **decisione 2005/671/GAI del Consiglio** sancisce che per combattere il terrorismo è fondamentale che tutti i servizi interessati possano disporre di informazioni il più possibile complete e aggiornate; onde le autorità competenti degli Stati membri devono trasmettere a Eurojust le informazioni concernenti le azioni penali o le condanne penali per reati di terrorismo che interessano o possono interessare due o più Stati membri. Assistere le autorità competenti degli Stati membri per assicurare un coordinamento ottimale delle indagini e delle azioni penali, compresa l'individuazione dei

(UE) 2023/2131 del Parlamento europeo e del Consiglio del 4 ottobre 2023 (che modifica il regolamento UE 2018/1727 del Parlamento europeo e del Consiglio e la decisione 2005/671/GAI del Consiglio), per quanto riguarda lo **scambio digitale di informazioni nei casi di terrorismo** ha innestato l'art. 21-*bis* nel regolamento UE 2018/1727 cit., disciplinando lo **scambio di informazioni sui casi di terrorismo** nei seguenti termini: «§ 1. Per quanto riguarda i reati di terrorismo, le autorità nazionali competenti informano i propri membri nazionali delle **indagini penali in corso** o concluse sotto il controllo di autorità giudiziarie non appena il caso è **deferito** alle autorità giudiziarie conformemente al diritto nazionale, in particolare il diritto processuale penale nazionale, di azioni penali e procedimenti giudiziari in corso o conclusi, e decisioni giudiziarie in merito a reati di terrorismo. Tale obbligo si applica a tutte le indagini penali riguardanti i reati di terrorismo, indipendentemente dal fatto che sussista un collegamento noto con un altro Stato membro o un paese terzo a meno che l'indagine penale, per le sue circostanze specifiche, interessi chiaramente un solo Stato membro. § 2. Il paragrafo 1 **non si applica** qualora: a) la condivisione di informazioni comprometta un'indagine in corso o la sicurezza di una persona; o b) la condivisione di informazioni sia in contrasto con gli interessi essenziali di sicurezza dello Stato membro interessato. § 3. Ai fini

---

collegamenti tra tali indagini e azioni penali, è un compito importante di Eurojust a norma del regolamento (UE) 2018/1727, che consente un approccio più proattivo e di fornire servizi migliori agli Stati membri, ad esempio suggerendo l'avvio di indagini e individuando esigenze di coordinamento, casi che potrebbero violare il principio ne bis in idem e lacune nell'azione penale. Nel settembre 2019 Eurojust ha istituito il registro giudiziario europeo antiterrorismo sulla base della e organizzazioni terroristiche sono sempre più coinvolte in altre forme gravi di criminalità e spesso fanno parte di reti organizzate. Tale coinvolgimento riguarda reati gravi quali la tratta di esseri umani, il traffico di stupefacenti, la criminalità finanziaria e il riciclaggio di denaro. È necessario **effettuare controlli incrociati sui procedimenti giudiziari contro tali forme gravi di criminalità**. Per consentire di individuare i collegamenti tra i procedimenti giudiziari transfrontalieri a carico di indagati per reati di terrorismo nonché i collegamenti tra i procedimenti giudiziari a carico di indagati per reati di terrorismo e le informazioni trattate relative ad altre forme gravi di criminalità, è fondamentale che Eurojust riceva dalle autorità nazionali competenti il più rapidamente possibile, le informazioni necessarie per poter individuare tali collegamenti mediante controlli incrociati decisione 2005/671/GAI con l'obiettivo specifico di individuare i potenziali collegamenti tra i procedimenti giudiziari a carico di indagati per reati di terrorismo e le eventuali esigenze di coordinamento derivanti da tali collegamenti. Tale scambio è ora disciplinato dal Regolamento (UE) 2023/2131 del Parlamento europeo e del Consiglio del 4 ottobre 2023 (che modifica il regolamento UE 2018/1727 del Parlamento europeo e del Consiglio e la decisione 2005/671/GAI del Consiglio), per quanto riguarda lo **scambio digitale di informazioni nei casi di terrorismo.** Le autorità nazionali competenti dovranno trasmettere le informazioni a Eurojust in modo strutturato, organizzato, sistematico e semiautomatizzato (che ricorre nei casi in cui la trasmissione delle informazioni è in parte automatizzata e in parte soggetta a controllo umano), prevedendosi che tale modo di trasmissione aumenterà in maniera significativa la qualità e la pertinenza delle informazioni ricevute da Eurojust. In particolare, la **condivisione, la conservazione e il controllo incrociato dei dat**i aumenteranno considerevolmente la quantità di dati trattati da Eurojust. Lo scambio di dati di identificazione affidabili è fondamentale affinché Eurojust individui i collegamenti tra le indagini in materia di terrorismo e i procedimenti giudiziari a carico di indagati per reati di terrorismo. È inoltre fondamentale affinché Eurojust crei e conservi un insieme di dati che garantisca un'identificazione affidabile delle persone oggetto di tali indagini o procedimenti giudiziari in materia di terrorismo. L'utilizzo dei dati biometrici è pertanto importante, tenuto conto delle incertezze relative ai dati alfanumerici, in particolare per i cittadini di paesi terzi, del fatto che talvolta gli indagati usano identità false o doppie e che tali dati biometrici sono spesso l'unico collegamento agli indagati nella fase delle indagini. Pertanto, qualora, a norma del diritto nazionale in materia di procedimenti penali o di diritti procedurali nei procedimenti penali, le autorità nazionali competenti conservino e raccolgano dati biometrici e siano autorizzate a trasmetterli, tali autorità dovrebbero poter scambiare tali dati, se disponibili, con Eurojust. Considerati la natura sensibile dei **dati biometrici** e l'impatto del trattamento di tali dati sul rispetto della vita privata e della vita familiare e sulla protezione dei dati di carattere personale, quali sanciti dagli articoli 7 e 8 della Carta dei diritti fondamentali dell'Unione europea, tali dati dovrebbero essere trasmessi in modo tale da rispettare rigorosamente i principi di necessità, proporzionalità e limitazione dello scopo, nonché per il fine esclusivo di identificazione di persone soggette a procedimenti penali connessi a reati di terrorismo. Poiché le informazioni sui collegamenti esistenti con altri procedimenti giudiziari sono maggiormente utili in una fase precoce dell'indagine, è necessario che le autorità nazionali competenti forniscano le informazioni a Eurojust non appena il caso è deferito a un'autorità giudiziaria conformemente al diritto nazionale. Un caso dovrebbe essere considerato deferito a un'autorità giudiziaria quando, ad esempio, l'autorità giudiziaria è informata di un'indagine in corso, autorizza o dispone una misura investigativa o decide di esercitare l'azione penale, a seconda del diritto nazionale applicabile. Se è già a conoscenza di collegamenti tra un procedimento penale nel suo Stato membro e un procedimento penale in un altro Stato membro, l'autorità nazionale competente dovrebbe informarne Eurojust.

del presente articolo, per reati di terrorismo si intendono i reati di cui alla direttiva (UE) 2017/541 del Parlamento europeo e del Consiglio. § 4. Le informazioni trasmesse a norma del paragrafo 1 includono i dati personali operativi e i dati non personali di cui all'allegato III. Tali informazioni possono includere dati personali a norma dell'allegato III, lettera d), ma solo se tali dati personali sono conservati dalle autorità nazionali competenti o possono essere comunicati alle stesse in conformità del diritto nazionale e se la trasmissione di tali dati è necessaria per identificare l'interessato in modo affidabile ai sensi dell'articolo 27, paragrafo 5. § 5. Fatto salvo il paragrafo 2, le autorità nazionali competenti informano i propri membri nazionali di qualsiasi modifica delle informazioni trasmesse a norma del paragrafo 1 senza indebito ritardo e, ove possibile, entro dieci giorni da tali modifiche. § 6. L'autorità nazionale competente non è tenuta a fornire tali informazioni se sono già state trasmesse a Eurojust. § 7. L'autorità nazionale competente può chiedere in qualsiasi fase l'assistenza di Eurojust nel seguito da dare relativamente ai collegamenti individuati sulla base delle informazioni fornite a norma del presente articolo».

In tal senso, può essere utile ripercorrere i **principali *tools* tecnologici** impiegati per contrastare il terrorismo internazionale, specie quello di matrice jihadista.

Il **riconoscimento facciale** (*facial recognition*) è uno strumento di scansione (**scanning)** impiegato in materia di immigrazione per le identificazioni alle frontiere[187], ai sensi del Regolamento (UE) 2018/1861 del Parlamento europeo e del Consiglio sull'istituzione, l'esercizio e l'uso del sistema d'informazione Schengen (SIS) nel settore delle verifiche di frontiera"[188]; per l'antiterrorismo, il riconoscimento facciale è in fase di test presso le forze dell'ordine di vari paesi; tale tecnologia consiste nell'estrazione ed elaborazione da parte dei sistemi di riconoscimento facciale di dati biometrici del volto di individui[189], raccolti mediante telecamere installate in luoghi pubblici, per raffrontarli con altri visi, raccolti in *databases*, di persone sospettate di terrorismo oppure già condannate per questo tipo di reati[190].

Altri algoritmi vengono impiegati per **analizzare i metadati**[191] **delle comunicazioni** che avvengono *online*, come gli indirizzi IP utilizzati ed i siti *web* visitati, per individuare comportamenti potenzialmente pericolosi ed eventualmente effettuare segnalazioni alle competenti autorità pubbliche[192]. In Francia, la *loi* 2015-912 permette ai servizi di *intelligence* di utilizzare questo tipo di algoritmi, detti *boîtes noires*, su autorizzazione del Primo Ministro, sentito il parere della *Commission Nationale de l'Informatique et des Libertés*, ancorché a prescindere da qualsiasi vaglio giurisdizionale di tipo preventivo[193].

---

[187] V. S. PENASA, *Migrazioni e intelligenza artificiale: nuovi percorsi di ricerca*, in *ADiM Blog*, dicembre 2021.

[188] *https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=celex:32018R1861*.

[189] K. HUSZTI-ORBÁN, *Use of Biometric Data to Identify Terrorists: Best Practice or Risky Business*? Human Rights Center, University of Minnesota, 2020.

[190] Si veda lo studio della FUNDAMENTAL RIGHTS AGENCY, *Facial Recognition Technology: Fundamental Rights Consideration in the Context of Law Enforcement*, 21 November 2019. La prima pronuncia giurisprudenziale sull'utilizzo di un sistema di riconoscimento facciale si è avuta da parte della High Court of Justice of England and Wales, che, nel settembre 2019, ha esaminato l'utilizzo, nell'ambito di grandi eventi sportivi, di un sistema di riconoscimento facciale da parte della polizia locale del Galles del Sud. *R(Bridges) v. The Chief Constable of South Wales Police et al.*, [2019] EWHC 2341. V., per un primo commento della decisione, A. PIN, *Non esiste la "pallottola d'argento": l'*Artificial Face Recognition *al vaglio giudiziario per la prima volta*, in *DPCE Online*, 4/2019, 3075 ss.

[191] I metadati permettono di conoscere significative informazioni relative alle comunicazioni, senza rivelarne il contenuto (si pensi alla durata di una chiamata telefonica, al luogo da cui viene inviata un'e-mail, all'orario in cui si accede ad un determinato sito, ecc.). Per approfondimenti cfr. E. GUILD-S. CARRERA, *The Political and Judicial Life of Metadata: Digital Rights Ireland and the Trail of the Data*, CEPS Liberty and Security in Europe Papers No.65 (May 29, 2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2445901 [Accessed October 20, 2014].

[192] Sul punto, F. TRÉGUER, *From Deep State Illegality to Law of the Land: The Case of Internet Surveillance in France*, in *Archive ouverte en Sciences de l' Homme et de la Société*,

[193] *Loi* n° 2015-912 *du* 24 *juillet* 2015 *relative au renseignement*. V. M-A. GRANGER, *Oversight of the State Emergency in France*, in B.J. GOOLD-L. LAZARUS (eds.), *Security and Human Rights*, Hart Publishing, Oxford, 2019,

La possibilità di avere a disposizione, incrociare e analizzare rapidamente, grazie alla c.d. **big data analytics** [194], un significativo numero di dati ha richiesto la predisposizione di strumenti giuridici per regolarne l'utilizzo, assieme ai meccanismi usati per processare le informazioni[195]. Sovente, sono emersi dubbi sulla compatibilità di questi strumenti con i diritti della persona, quali quelli della *privacy*, della *data protection*, ma anche con principi fondamentali, come la presunzione di innocenza; le pronunce dei giudici non hanno mancato di censurare specifici aspetti delle corrispondenti normative sull'impiego delle tecniche di sorveglianza[196].

Le **attività di sorveglianza** hanno assunto un carattere massivo, con sviluppo di nuove tecnologie come la **criptoanalisi** (*signals intelligence*)[197], nonché la predisposizione di ***databases*** in grado di **conservare**[198] **e confrontare i c.d. metadati**, utili per la profilazione degli individui con scopo criminal-preventivo.[199] Sul punto va registrato che sovente le Corti hanno rilevato la violazione dei limiti minimi di tutela dei diritti fissati dai cataloghi costituzionali o sovranazionali; quando non hanno escluso la possibilità teorica della legittimità della misura antiterrorismo sottoposta al loro scrutinio, ad esempio affrontando la questione della legittimità della sorveglianza di massa[200], ne hanno comunque riscontrato la non conformità, sul piano pratico, alle garanzie dei diritti tutelati dalle fonti sovranazionali o costituzionali. Come è stato in proposito rilevato, «si può dunque evidenziare il *gap* fra la dimensione teorica e quella pratica. In termini più chiari, benché l'azione antiterroristica, messa a punto dai legislatori per salvaguardare la sicurezza (o almeno la sua percezione sociale), venga, sul piano teorico-astratto, ritenuta compatibile con il quadro normativo di riferimento, sul piano pratico, non supera però il vaglio delle corti, che, in qualche caso, si spingono a formulare una serie di linee guida utili per la necessaria (successiva) opera di revisione delle misure da applicare in concreto»[201].

---

389 ss.

194 Per funzionare gli algoritmi hanno necessità di un'enorme quantità di dati. Sul rapporto fra algoritmi e *big data*, v. P. COSTANZO, *La democrazia digitale (precauzioni per l'uso)*, in *Diritto pubblico*, 1/2019, 71 ss.

195 Sul punto, E. GUILD-S. CARRERA, *The Political and Judicial Life of Metadata: Digital Rights Ireland and the Trail of the Data Retention Directive*, CEPS Paper in Liberty and Security, CEPS, Brussels, 2014.

196 Cfr. per lo scrutinio sulle tecniche di sorveglianza di massa, tra le decisioni più importanti, si vedano: Corte di giustizia dell'Unione europea, sent. 8 aprile 2014, C-293/12 e C-594/12, *Digital Rights Ireland*; Id., sent. 6 ottobre 2015, C-362/14, *Schrems*; Id., sent. 21 dicembre 2016, C-203/15 e C-698/15, *Tele2 Sverige AB*.

197 Consiste nella raccolta di informazioni grazie all'intercettazione e all'analisi di comunicazioni cifrate che vengono decrittate; Cfr. M.V. HAYDEN, *Balancing Security and Liberty: The Challenge of Sharing Foreign Signals Intelligence*, in *Notre Dame Journal of Law, Ethics and Public Policy*, 2005, 247 ss.

198 Ai sensi dell'**art. 132, comma 5-bis d.lgs. n. 196/2003** (introdotto dall'art. 24 L. 20/11/2017, n. 167) in attuazione dell'articolo 20 della direttiva (UE) 2017/541 del Parlamento europeo e del Consiglio, del 15 marzo 2017, sulla lotta contro il terrorismo e che sostituisce la decisione quadro 2002/475/GAI del Consiglio, al fine di garantire strumenti di indagine efficace in considerazione delle straordinarie esigenze di contrasto del terrorismo, anche internazionale, per le finalità dell'accertamento e della repressione **dei reati di cui agli articoli 51, comma 3-quater, e 407, comma 2, lettera a), del codice di procedura penale il termine di conservazione dei dati di traffico telefonico e telematico nonché dei dati relativi alle chiamate senza risposta**, di cui all'articolo 4-bis, commi 1 e 2, del decreto-legge 18 febbraio 2015, n. 7, convertito, con modificazioni, dalla legge 17 aprile 2015, n. 43, è stabilito **in settantadue mesi, in deroga a quanto previsto dall'articolo 132, commi 1 e 1-bis, del codice in materia di protezione dei dati personali**, di cui al decreto legislativo 30 giugno 2003, n. 196.

199

200 Cfr. il Parere 1/15 della **Corte di giustizia dell'Unione europea**, con cui pur legittimando, a livello teorico, la sorveglianza di massa come misura potenzialmente adatta a prevenire attacchi terroristici e, quindi, a tutelare la pubblica sicurezza, è censurato però l'accordo internazionale stretto tra Canada e Unione europea sullo scambio di dati riguardati il traffico aereo; cfr. Corte di giustizia dell'Unione europea, A-1/15, 26 luglio 2017, in tema A. VEDASCHI, *L'Accordo internazionale sui dati dei passeggeri aviotrasportati (PNR) alla luce delle indicazioni della Corte di giustizia dell'Unione europea*, in *Giur. cost.*, 4/2017, 1913 ss.

201 A. VEDASCHI, *Intelligenza artificiale e misure antiterrorismo alla prova del diritto costituzionale*, in *Consulta Online*, 17 febbraio 2020; cfr. il Parere 1/15 cit., nota 9, spec. §§ 154-231.

- **6.5.5. In particolare, il contrasto dell'IA alla radicalizzazione terroristica**.

Tra gli ambiti di impiego della IA nella lotta al terrorismo di rilievo appare l'azione preventiva contro la **radicalizzazione**[202]. Parte della dottrina ha provato a **definire** la radicalizzazione come il processo attraverso il quale gli individui possono diventare estremisti violenti»[203]; il concetto di «violent extremism» è a sua volta definito dalle Nazioni Unite come «le convinzioni e le azioni di persone che sostengono o usano la violenza per raggiungere obiettivi ideologici, religiosi o politici», includendo «terrorismo e altre forme di violenza motivata da ragioni politiche»[204].

Tra le forme di radicalizzazione distinguibili in termini fenomenologici, prima che strettamente giuridici (per le quali si rinvia al § 5.2) possono essere annoverate: i) l'indottrinamento[205], suscettibile di sfociare nel proselitismo e il reclutamento in organizzazioni terroristiche ed usualmente realizzata nel *dark web*; ii) l'apprezzamento o l'elogio nei confronti delle ideologie estremiste oppure di azioni terroristiche compiute nel passato, espressi anche grazie all'utilizzo dei principali *social networks e* che pone seri problemi di confine tra libera manifestazione del pensiero e affermazioni da reprimere*; iii)* l'umiliazione delle vittime di attentati terroristici, attuata su piattaforme online[206]; iv) l'autoconvincimento individuale (c.d. *lone wolf*) con adesione a idee estremiste e violente, all'esito, invero, sovente di messaggi lasciati da altre persone.

Tra i metodi di contrasto della radicalizzazione il principale è l'**identificazione e la rimozione** del messaggio "pericoloso" in quanto potenzialmente radicalizzante[207]. Come anticipato, il "discorso terroristico" viene diffuso sempre più spesso con mezzi informatici, ossia utilizzando il *web* (sia *dark web* sia *surface web*, a seconda di quanto "esplicite" siano le affermazioni)[208]. Di conseguenza, si pone la necessità di scandagliare una rete amplissima, difficilmente controllabile dall'occhio umano. Per questo motivo, le piattaforme *online* ricorrono in modo massiccio a strumenti automatizzati (c.d. algoritmi intelligenti), in grado di rilevare il rischio insito nei contenuti di Internet e apporre un *flag* agli stessi, segnalandoli come potenzialmente dannosi per la sicurezza in quanto probabilmente radicalizzanti.

Dal punto di vista **tecnologico, gli strumenti** più utilizzati dai *providers* di servizi *online* per **identificare i messaggi potenzialmente radicalizzanti e disporne l'eliminazione** sono quattro[209]:

---

[202] Sulla complessità definitoria del concetto di "radicalizzazione", cfr. C. GRAZIANI, *Terrorismo internazionale, radicalizzazione e tecnologia*, in *Federalismi*, 31 maggio 2023.

[203] P. ROMANIUK, *Does CVE Work? Lessons Learned from the Global Effort to Counter Violent Extremism. Washington, DC: Global Center on Cooperative Security*, 2015.

[204] United Nations, *Preventing violent extremism through education: A guide for policy-makers*, United Nations Educational, Scientific and Cultural Organization, Paris, 2017

[205] Si tratta di incitamenti espliciti ad abbracciare l'ideologia terrorista, spesso anche con indicazioni dettagliate e "pratiche" circa le modalità per portare avanti la causa jihadista (ad esempio, indicazione circa la costruzione di armi, l'organizzazione di veri e propri "viaggi" per poter combattere nelle file dello Stato Islamico, ecc.).

[206] Alcuni ordinamenti hanno previsto specifiche fattispecie di reato per questo tipo di condotta, distinte sia dai crimini di apologia sia da quelli che puniscono l'offesa nei confronti della vittima e dei suoi familiari. Con particolare riferimento al caso spagnolo, vi è il reato di *enaltecimiento* (art. 578 c.p.), che è integrato da condotte elogiative di atti o ideologie terroristiche ed è inserito nel capo del codice penale dedicato ai "delitti di terrorismo", dunque il bene messo in pericolo è la sicurezza pubblica; inoltre, vi è il reato di *humillación* delle vittime di terrorismo, che consiste in affermazioni offensive e sprezzanti ai danni delle stesse; lo stesso comportamento integra il reato di *injuria* (art. 208 c.p.), che si trova invece nel capo dedicato ai delitti contro l'onore e il cui bene giuridico leso è la dignità della persona. Nel caso in cui vi siano comportamenti offensivi di vittime di terrorismo, i giudici spagnoli tendono a cumulare le sanzioni per i due reati.

[207] Parte della dottrina americana ritiene che la rimozione del messaggio pericoloso potrebbe essere inefficiente, perché renderebbe più difficile identificare "covi" di terroristi e agire per perseguirli. Cfr. M. LAVI, *Do Platforms Kill?* in *Harvard Journal of Law and Public Policy*, vol. 43, n. 2, 2020, pp. 477-573.

[208] Per alcuni dati, cfr. ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *Transparency Reporting on Terrorist and Violent Extremist Content Online*, OECD Digital Economics Paper, July 2021.

[209] S. MCDONALD, S. GIRO CORREIA, A.L. WATKIN, *Regulating Terrorist Content on Social Media: Automation and the Rule of Law*, in *International Journal of Law in Context*, vol. 15, 2019, pp. 183-197. V. anche A. ZORNETTA, I. POHLAND, *Legal and Technical Trade-Offs in the Content Moderation of Terrorist Live-Streaming*, in *International Journal of Law and Information Technology*, vol. 30, 2022, pp. 302-330

il *natural language processing* (NLP), l'*image matching*, il *clustering* e gli algoritmi di *recidivism tackling* (che si basano in parte sul *clustering*).

Le tecniche di **elaborazione del linguaggio naturale** (*natural language processing)* **utilizzano** algoritmi "nutriti" con i testi di messaggi in precedenza rimossi perché ritenuti pericolosi, in modo da istruirli nell'apprendimento di parole e combinazioni di parole "rischiose". Esistono tre principali applicazioni delle tecniche di NLP al discorso terroristico[210]: l'analisi *lexical or vectorial based* [211], il *neural language modell*[212]; le *syntactic and semantic features*[213]. Le tecniche di NLP descritte funzionano sia su documenti testuali (ad esempio, un *tweet* pubblicato in una data lingua) sia su files audio, se impiegato un programma di trascrizione per estrarne una versione scritta. In generale, l'uso del NLP per identificare contenuti terroristici si basa sull'assunto che la macchina (ossia l'IA) possa imparare a "leggere" i messaggi non solo a livello di riconoscimento dei grafemi e dei loro singoli significati, ma anche valutando la combinazione delle parole e il contesto in cui inserite. Al fine di potenziare la valutazione del contesto, spesso il NLP viene combinato con le altre strategie descritte di seguito.

Le tecniche di **riconoscimento di immagini** (*image matching*)[214] allenano gli algoritmi sulla base delle immagini (siano esse statiche o video) che possono indurre la radicalizzazione in taluni individui. In questo modo, l'IA impara a riconoscerle se si imbatte in immagini uguali sul *web*. L'*image matching* si basa in larga parte su contenuti terroristici in precedenza identificati e rimossi, forniti come *input* all'algoritmo affinché possa "imparare" come appare un'immagine o un video di matrice terroristica ed essere in grado di rilevare tali contenuti scandagliando il *web*. L''*image matching* migliora e si raffina grazie all'esperienza: quanti più contenuti terroristici vengono individuati (manualmente o grazie a precedenti algoritmi), tanto più il *set* di dati da inserire nella programmazione dell'algoritmo sarà completo e vario, e dunque porterà tale algoritmo ad essere maggiormente funzionale. Per potenziare questo tipo di algoritmi di *image matching*, il *Global Internet Forum to Counter Terrorism* (GIFCT)[215] ha elaborato e continua ad alimentare il c.d.

---

[210] J. TORREGROSA et al., *A survey on extremism analysis using natural language processing: definitions, literature review, trends and challenges*, in J*ournal of Ambient Intelligence and Humanized Computing*, 2022.

[211] Con tale applicazione un testo presente in rete viene rappresentato come un insieme di "punti" in uno spazio, detto *vectoral space model* (VSM); ogni punto viene denominato *token* ed è composto da un'aggregazione di termini lessicali cui viene assegnata una certa percentuale di rischio. Costruito il VSM, l'algoritmo analizza il rischio complessivo insito nello stesso, sulla base di indicatori quali il ricorrere di termini o accostamenti di termini a rischio particolarmente alto, l'assenza di taluni termini visti come rischiosi, il ricorrere di due o più *token a priori* considerati a rischio insieme. In questo modo, si tenta di rendere l'algoritmo capace di analizzare il contesto, oltre che le singole parole: l'espressione "cellula terroristica", potenzialmente ad alto rischio, potrebbe essere frequente in un covo *online* di terroristi che si esprime via *chat*, come in uno scritto accademico che si occupi di radicalizzazione; un siffatto scritto, però, non vedrà una frequenza altrettanto alta di *altre* espressioni a rischio (come incitamenti all'adesione). Pertanto, l'algoritmo dovrebbe unire i due fattori (alta ricorrenza del sintagma "cellula terroristica", ma ricorrenza nulla di parole che esprimano un invito ad aderirvi) e, di conseguenza, essere in grado di non *flaggare* come pericoloso lo scritto accademico.

[212] Sulla base di questa tecnica, all'algoritmo vengono fornite **informazioni sulle relazioni esistenti fra parole**, che poi applica nell'analizzare i testi rinvenuti su internet. Per esempio, la parola "terrorismo" non è necessariamente indice di un messaggio radicalizzante, ma è probabile che lo sia se abbinata ad altri termini o se presente in un ambiente con date caratteristiche. Pertanto, all'algoritmo viene istruito oltre che a riconoscere il termine "principale", il rischio che deriva da ciascun termine "abbinato", in modo che il sistema possa appore un *flag* in caso di percentuali di rischio particolarmente alte.

[213] Si tratta di algoritmi capaci di rilevare la portata emozionale di una affermazione, identificando il tipo di sentimento espresso da una frase o da un gruppo di frasi (in alcuni casi, si parla di *sentiment analysis*). Si tratta di tecniche molto avanzate e soggette a significative possibilità di errore, per inesatta valutazione del contesto o funzionamento non perfetto, su una macchina di una valutazione assai complessa, quella dei sentimenti e delle emozioni che possono celarsi dietro uno scritto, potenzialmente soggetta a equivoci anche quando attuata dall'essere umano.

[214] Cfr. D. GARCÍA-RETUERTA et al., *Counter-Terrorism Video Analysis Using Hash-Based Algorithms*, in *Algorithms*, vol. 12, 2019, pp. 110-119

[215] Il Global Internet Forum to Counter Terrorism è una *partnership* volontaria tra i maggiori operatori del *web* (ad esempio, Google/YouTube, Microsoft, Twitter, i quali sono fra i fondatori).

database of hashes[216]. Alcune difficoltà si potrebbero generare quando l'autore di un contenuto lo modifica leggermente, variando l'*hash*. In questi casi, esistono algoritmi programmati per identificare anche il contenuto *simile* anche se la percentuale di errori che l'AI può commettere si alza. Inoltre, sempre in caso di contenuto simile, l'*image matching* aumenta le proprie possibilità di successo se abbinato ad un algoritmo di NLP che può esaminare al contempo eventuali contenuti audio (a condizione che essi consistano in parole e non, ad esempio, in meri contenuti musicali), previa trascrizione con un programma di scrittura automatico.

Nella tecnica del **raggruppamento** (*clustering*)[217] un algoritmo riunisce oggetti con caratteristiche simili estraendo i caratteri comuni a più profili di soggetti condannati o segnalati dalle autorità come terroristi. Dato un certo numero di *account* "di partenza" (necessari per programmare l'algoritmo e "insegnargli" a riconoscere i profili pericolosi), l'algoritmo può estrarre anche più di un *cluster*[218]. Qualora lo stesso *pattern* (modello, o insieme di caratteristiche) si ritrovi in altri profili di utenti della rete l'algoritmo è in grado di misurare il livello di compatibilità con il *cluster* identificato e, se tale livello è alto, segnalare il profilo. Le caratteristiche rilevanti, che vanno a comporre il *pattern di* un dato *cluster*, possono essere di diversi tipi: *i*) la manifestazione di apprezzamento per determinate pagine o contenuti; ii) l'intrattenere una rete di contatti (tramite l'opzione *follow* sui *social networks*) con altre persone, a loro volta sospettate; iii) il frequente *re-post* di contenuti di alcuni siti o canali[219].

Gli **algoritmi per recidivi** (*repeat offenders algorithms*) [220], infine, valutano se i nuovi *accounts* che accedono ad una data piattaforma costituiscano profili *fake* di soggetti la cui utenza era stata già disabilitata per aver pubblicato contenuti terroristici. Tale tecnica si basa su una combinazione delle strategie descritte in precedenza. In particolare, si può usare l'analisi dei *patterns* e la comparazione con *patterns* dei profili precedentemente disabilitati, l'identificazione dell'indirizzo IP, del posizionamento geografico, della rete di contatti riconducendo il nuovo profilo ad un dato *cluster*, il riconoscimento di immagini che, seppur disabilitate dopo essere state caricate con una data utenza, potrebbero essere reinserite utilizzando il nuovo *account*. Questa tipologia di algoritmi apre al pericolo che la suddivisione in *clusters*, operata per identificazione i recidivi, finisca per avere effetti discriminatori. Infatti, l'algoritmo "impara" non solo dall'*input* originario, inserito in fase di programmazione, ma anche dall'esperienza di funzionamento; perciò, se in concreto ha evidenziato che i profili maggiormente pericolosi sono localizzati in una data area geografica, oppure hanno determinate caratteristiche che riportano ad un certo credo religioso o provenienza etnica, potrebbe "attenzionare" maggiormente i profili in cui alcuni di questi caratteri ricorrono, sottoponendoli ad

---

[216] Gli *hashes* sono codici identificativi univoci, paragonabili ad una sorta di "impronta digitale", assegnati a ciascun contenuto multimediale, che lo rende immediatamente riconoscibile. Per esemplificare il funzionamento del *database of hashes*, si può immaginare un video pericoloso (perché incitante al terrorismo) che sia stato caricato su YouTube da un utente radicalizzato e poi rimosso da questa piattaforma. Il proprietario del video potrebbe a quel punto caricarlo su altra piattaforma, per esempio Facebook (ora Meta). Se, però, YouTube ha condiviso l'*hash* di quel video grazie al *database of hashes*, Facebook (Meta) potrà prontamente rimuoverlo, perché identificherà autonomamente che l'*hash* del nuovo *upload* coincide con un *hash* già segnalato.

[217] G.M. CAMPEDELLI et al., *A complex networks approach to find latent clusters of terrorist groups*, in *Applied Network Science*, Vol. 4, 2019, pp. 59-81.

[218] Per esempio, si potrebbe avere il *cluster* A, con certe caratteristiche che corrispondono ai profili dei terroristi già condannati, il *cluster* B, con altre caratteristiche, in parte comuni al *cluster* A ma in parte differenti, che coincidono con i soggetti solo sospettati di terrorismo; e così via.

[219] I modelli nascosti (*hidden patterns*) sono i più complessi, trattandosi di somiglianze che si riscontrano in gruppi di soggetti diversi dal punto di vista ideologico, culturale e geografico (per cui l'algoritmo difficilmente potrebbe evidenziare identità nell'espressione di *follow* o nella rete dei contatti) e consistono, ad esempio, in parole chiave simili utilizzate nelle stringhe di ricerca (aspetto opaco, poiché la ricerca effettuata *online* lascia una traccia meno chiara rispetto alla pubblicazione di un *post* o all'espressione di un *like* rispetto ad un certo contenuto).

[220] V. CROSSET, B. DUPONT, *Cognitive Assemblages: The Entangled Nature of Algorithmic Content Moderation*, in *Big Data & Society*, vol. 9, n. 2, 2022, pp. 1-13

una sorta di "sorveglianza rinforzata"[221]. Affidando agli algoritmi la fase dell'identificazione degli atti di terrorismo, si pone un rischio di moltiplicazione esponenziale di possibili pregiudizi (*biases*) potrebbero contenere, ma difficile da rilevare, se si considera che essi sono coperti dalle regole in materia di *intellectual property* e conoscibili solo dall'azienda che li usa.

Le tecnologie innovative possono essere utilizzate per identificare coloro che sono più vulnerabili ed esposti agli sforzi radicalizzanti dei gruppi terroristici. Il **metodo di reindirizzamento** (*Redirect method*) [222] consiste in un sistema algoritmico capace di individuare i profili degli individui potenziali obiettivi della narrativa dell'odio. L'algoritmo viene testato sul discorso islamista, allenato a riconoscere i vocaboli e le frasi utilizzati dalla retorica jihadista: una volta programmato, viene usato per intercettare l'utente di internet sensibile al discorso terroristico, con l'obiettivo di "re-indirizzarlo" verso pagine *web* di contro-narrativa[223], diminuendo le possibilità di adesione a ideologie violente o estremiste.

Ulteriori algoritmi vengono impiegati per l'**individuazione e la cancellazione dei messaggi a sfondo terroristico** trasmessi attraverso il *web*. *Service providers*, *hosting providers*, *social networks* e motori di ricerca stanno investendo risorse e ricerca in apparati tecnologici che possano scandagliare la rete alla ricerca di messaggi con contenuti terroristici, violenti o pericolosi, per rimuoverli, in ottemperanza alle policies interne e ad atti normativi[224]. È indubbio che, rispetti ai *media* tradizionali, i moderni *social media* consentono ai terroristi di mantenere il controllo diretto dell'immagine che proiettano, producendo, diffondendo e propagandando i propri contenuti su scala globale, nella "*mass self-communication"*[225].

Rispetto alla **rimozione** dei contenuti terroristici sul *web*, sino al 2015, la **regolamentazione** pubblica internazionale si è mostrata prudente, favorendo iniziative di partenariato pubblico-privato. Alcune risoluzioni del Consiglio di Sicurezza hanno esaminato i rischi posti dallo sfruttamento di Internet da parte delle organizzazioni terroristiche, il cosiddetto *e-terrorism*[226] senza citare strumenti automatizzati o algoritmi di IA, né porsi quale quadro giuridico di cui gli Stati potessero servirsi per la rimozione dei contenuti terroristici *online*, limitandosi a richiamare accordi, informali e su base volontaria, fra le più importanti aziende tecnologiche (Facebook, Google per quanto riguarda YouTube[227]) , impegnate ad evitare, anche mediante l'ausilio di sistemi automatizzati, disciplinati nelle *policies*, che sui loro spazi virtuali fossero presenti contenuti pericolosi. In questa prima fase,

---

[221] Proprio su questo tema vi è stato il caso *Federal Bureau of Investigation v. Fazaga*, 595 U.S. (2022) in cui un gruppo di musulmani sosteneva di essere stato sottoposto ad una sorveglianza automatizzata "speciale" (ossia non attuata nei confronti di altri cittadini) solo in ragione del loro credo religioso. La questione, giunta all'attenzione della Corte Suprema federale, non è stata esaminata da un punto di vista sostanziale, poiché la Corte ha ritenuto che, per valutare le motivazioni della sorveglianza, si sarebbe dovuti entrare nel merito della stessa e dei dati a disposizione del FBI, ma, trattandosi di informazioni coperte da *state secrets privilege* invocato dall'Esecutivo, ciò non era possibile per un organo giurisdizionale.

[222] Pilotato da Jigsaw e Moonshot nel 2016 e successivamente distribuito a livello internazionale da Moonshot in collaborazione con aziende tecnologiche, governi e organizzazioni di base.

[223] A. VEDASCHI, *Intelligenza artificiale e misure antiterrorismo alla prova del diritto costituzionale*, in *Consulta Online*, 17 febbraio 2020.

[224] A. VEDASCHI, *Intelligenza artificiale e misure antiterrorismo alla prova del diritto costituzionale*, in *Consulta Online*, 17 febbraio 2020.

[225] B. NACOS, *Mass-Mediated Terrorism: Mainstream and Digital Media in Terrorism and Counterterrorism. Third edition*; *Rowman & Littlefield eds*, 2016.

[226] Consiglio di Sicurezza delle Nazioni Unite, risoluzione n. 21129 del 17 dicembre 2013; Id., risoluzione n. 2354 del 24 maggio 2017; Id., risoluzioni nn. 2395 e 2396 del 21 dicembre 2017.

[227] Accordi *informali sono stati conclusi* tra i c.d. giganti del *web* (i principali motori di ricerca, come Google, e le relative piattaforme di condivisione, come YouTube, ma anche soggetti quali Twitter o Microsoft) che si sono impegnati ai fini di identificazione di contenuti potenzialmente radicalizzanti. In alcune circostanze, questi accordi coinvolgono anche le autorità pubbliche: è il caso dell'*EU Internet Forum to Counter Terrorism*, in cui la Commissione europea si è fatta capofila di un'intesa che include, fra gli altri, Google, Meta, Microsoft, oltre che la Radicalisation Awareness Network (RAN) e lo stesso GIFCT (Global Internet Forum to Counter Terrorism) che è una *partnership* volontaria tra i maggiori operatori del *web* (ad esempio, Google/YouTube, Microsoft, Twitter).

a livello di *soft law*, l'intervento del **settore privato** è risultato predominante, anticipando gli sforzi della regolazione istituzionale[228], senza pregiudicare l'essenza del fondamentale diritto alla libera manifestazione del pensiero[229].

I soggetti coinvolti appartengono a tre categorie: *i*) i c.d. ISP, ***Internet service providers***, organizzazioni o infrastrutture che offrono agli utenti, previa stipulazione di un contratto di fornitura, servizi inerenti a Internet; ii) gli ***hosting providers***, piattaforme *online* (ad es. Youtube) che "ospitano" contenuti di vario genere; iii) ICT ***(Information and Communication Technology companies)***, società di tecnologia dell'informazione e della comunicazione che consentono al fruitore sia di pubblicare contenuti sia di comunicare con altri utenti *online*. Tali operatori hanno offerto definizioni domestiche, chiarendo le caratteristiche del messaggio terroristico o radicalizzante[230] e comunicando le conseguenze dell'eventuale caricamento o condivisione in rete di contenuti corrispondenti. Ad esempio, YouTube ha chiarito che laddove l'algoritmo impiegato avesse individuato contenuti trasgressivi delle *policies*, gli stessi sarebbero stati eliminati e al soggetto che li aveva caricati sarebbe stata inviata un'*e-mail*. Ove la violazione si fosse ripetuta, l'*account* YouTube sarebbe stato disabilitato. Tralasciando profili afferenti alla tutela del titolare dell'account, le *policies* non richiamano espressamente l'utilizzo degli algoritmi e i casi in cui il mezzo automatico può operare. È noto, però, che la piattaforma Facebook usa l'IA per calcolare la probabilità che un *post* contenga un messaggio che supporta il terrorismo e in presenza di probabilità elevata procede ad una rimozione automatizzata, senza intervento dell'operatore umano[231]. Si tratta di norme adottate da soggetti privati mediante i *terms of services*, contenenti i c.d. *community standards*[232], documenti predisposti dagli organi delle singole società, con la collaborazione di esperti del settore. Questi atti, non riconducibili alle fonti ordinarie, hanno colmato inizialmente un vuoto normativo[233] lasciato dai regolatori pubblici; la scelta dei meccanismi tecnologici da impiegare è stata rimessa agli attori privati, che hanno avuto modo di affinare le proprie tecniche algoritmiche e di definire intese con autorità pubbliche[234], nell'ambito di una complessiva privatizzazione delle attività tradizionalmente pubbliche[235] (sotto i profili definitori delle nozioni di contenuti terroristici, del potere esecutivo di rimozione, con significativi limiti di trasparenza e di pubblicità degli algoritmi, e sin anche del potere giurisdizionale[236]). Il ruolo fondamentale giocato dalle linee guida e dalle *policies* interne adottate dagli attori digitali, con la funzione para-normativa affidata di fatto ai privati, si è assistito ad una

---

[228] C. GRAZIANI, *Intelligenza artificiale e fonti del diritto: verso un nuovo concetto di soft law? La rimozione dei contenuti terroristici online come case-study*, in *DPCE Online*, [S.l.], v. 50, n. Sp, mar. 2022.

[229] V. *Evolving an Institution*, in Global Internet Forum to Counter Terrorism, *GIFCT*.

[230] Tra le norme della *community* YouTube, compare il divieto di "*Contenuti prodotti da organizzazioni criminali violente o terroristiche […] che elogiano o commemorano figure di spicco degli ambienti terroristici e criminali allo scopo di incoraggiare altri a commettere atti di violenza*".

[231] Facebook, *Hard Questions: Questions: What Are We Doing to Stay Ahead of Terrorists? 2018, about.fb.com/news/2018/11/staying-ahead-of-terrorists/.*

[232] M. BORELLI, SOCIAL *media corporations as actors of counter-terrorism;* in New *Media and Society, 2021*.

[233] Ibidem.

[234] Nel Regno Unito specifici corpi di polizia (*Counter Terrorism Internet Referral Unit – CTIRU*) lavorano in collaborazione con i gestori di motori di ricerca e di piattaforme digitali, nell'intento di individuare il "rischio radicalizzazione"; Sul tema, si veda C. WALKER, *Blackstone's Guide to the Anti-Terrorism Legislation*, Oxford University Press, Oxford, 2014.

[235] Per una panoramica assai completa cfr. A. VEDASCHI, *Intelligenza artificiale e misure antiterrorismo alla prova del diritto costituzionale*; in *Consulta Online*, 17 febbraio 2020, che rileva come si stia assistendo ad una «torsione dell'idea di sovranità, che cessa di configurarsi come concetto riconducibile *in toto* allo spazio pubblico, allo Stato, per diventare invece oggetto di una parziale "condivisione" tra pubblico e privato, se non di una vera e propria "cessione" da parte delle autorità statali a favore del settore ICT».

[236] Di recente, Facebook ha annunciato l'imminente creazione di un *Oversight Board*, che dovrebbe entrare in funzione dai primi mesi del 2020. L'*Oversight Board* dovrà pronunciarsi sulle doglianze di utenti che lamentano l'ingiusta eliminazione di contenuti postati sulla rete; cfr. *post* di M. Zuckerberg, *A blueprint for content governance and enforcement*, 15 November 2018. L'*Oversight Board* è stato poi appellato "Facebook Supreme Court". Per una riflessione sui rischi derivanti dalla sua introduzione Q. WEINZIERL, *Difficult Times Ahead for the Facebook, Supreme Court*, in *Verfassungsblog*, 21 September 2019

"*soft law* dei privati"[237]

Dal 2015, si è aperta una seconda fase volta a formalizzare questi accordi (impostati secondo il modello della *public-private partnership*) in strumenti giuridici vincolanti. Dopo alcuni approcci a livello nazionale da parte di Francia e Germania[238], su stimolo di questi due Paesi l'Unione europea si è dotata del Regolamento (UE) 2021/784[239], applicabile dal 7 giugno 2022. Il testo normativo ha svolto uniformato le regole degli Stati membri in tema di rimozione dal *web* dei messaggi radicalizzanti, rispondendo alla preoccupazione che lo stesso possa diventare un «catalizzatore della radicalizzazione degli individui che può portare a atti terroristici»[240]; pertanto, i prestatori di servizi *online* sono tenuti a collaborare attivamente con le autorità per identificare l'eventuale uso improprio delle loro piattaforme. Nondimeno viene sottolineato il ruolo chiave della libertà di espressione, per evitare che «le opinioni radicali, polemiche o controverse espresse nell'ambito di dibattiti politici sensibili»[241] vengano considerate alla stregua di contenuti terroristici. Il regolamento considera "*terrorist content*" non solo le "istruzioni" per la commissione di reati di terrorismo o costruzione di armi e ordigni finalizzati allo stesso scopo, ma anche qualsiasi messaggio che "istighi" o "solleciti" la commissione di un reato terroristico; profilo non esente da vaghezza. Il regolamento non dettaglia gli strumenti tecnologici, come le tipologie di algoritmi, da utilizzare per la rimozione, lasciandole alla discrezione degli operatori del *web*. L'entrata in vigore del regolamento non elimina le *partnership* di taglio volontaristico tra operatori del *web*, che continuano ad esistere e, anzi, ad essere funzionali all'adozione delle c.d. misure proattive. L'**art. 2, punto 7 del Regolamento** (UE) 2021/784 definisce i **«contenuti terroristici**», quali « materiali che: a) istigano alla commissione di uno dei reati di cui all'articolo 3, paragrafo 1, lettere da a) a i), della direttiva (UE) 2017/541, se tali materiali, direttamente o indirettamente, ad esempio mediante l'apologia di atti terroristici, incitano a compiere reati di terrorismo, generando in tal modo il pericolo che uno o più di tali reati siano commessi; b) sollecita una persona o un gruppo di persone a commettere o a contribuire a commettere uno dei reati di cui all'articolo 3, paragrafo 1, lettere da a) a i), della direttiva (UE) 2017/541; c) sollecita una persona o un gruppo di persone a partecipare alle attività di un gruppo terroristico, ai sensi dell'articolo 4, lettera b), della direttiva (UE) 2017/541; d) impartisca istruzioni per la fabbricazione o l'uso di esplosivi, armi da fuoco o altre armi o sostanze nocive o pericolose, ovvero altri metodi o tecniche specifici allo scopo di commettere o contribuire alla commissione di uno dei reati di terrorismo di cui all'articolo 3, paragrafo 1, lettere da a) a i), della direttiva (UE) 2017/541; e) costituisca una minaccia di commissione di uno dei reati di cui all'articolo 3, paragrafo 1, lettere da a) a i), della direttiva (UE) 2017/541»[242].L'ordine di rimozione da parte dell'Autorità competente deve contenere, tra l'altro, una **la motivazione,** sufficientemente dettagliata, delle ragioni per cui i contenuti sono considerati contenuti terroristici e un riferimento alle pertinenti tipologie di materiale di cui all'articolo 2, punto 7.

---

[237] Sull'uso della *soft law* nell'ambito delle misure antiterrorismo, si veda ONU, Assemblea Generale, *Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism*, 29.8.2019, A/74/335.

[238] C. GRAZIANI, *Intelligenza artificiale e fonti del diritto: verso un nuovo concetto di* soft law*? La rimozione dei contenuti terroristici* online *come* case-study, in *DPCE Online*, n. speciale, 2022, pp. 1473-1490.

[239] Regolamento (UE) 2021/784 del Parlamento europeo e del Consiglio, del 29 aprile 2021, relativo al contrasto della diffusione di contenuti terroristici online, G.U.U.E. 172, 17.5.2021.

[240] Regolamento (UE) 2021/784, cit., considerando n. 5.

[241] Regolamento (UE) 2021/784, cit., considerando n. 12.

[242] In senso critico A. CISTERNA, *Un modello di contrasto efficace anche per le altre minacce in rete*, in Guida al Diritto n, 38, p. 41-42, il quale osserva che «mentre per alcune fattispecie (si veda lettera a ovvero *d)* la legislazione nazionale prevede espressa ipotesi di reato, la nozione di «minaccia di commissione)) (lettera *e)* rinvia a qualcosa di molto meno degli atti preparatori ex articolo 56 Cp e si colloca in un perimetro di estrema imprecisione e ampiezza. Se per altri Paesi la questione può assumere un rilievo neutro, la circostanza **che in Italia sia coinvolta l'autorità giudiziaria nell'applicazione di categoria di tale ampiezza, ai limiti dei poteri di sicurezza pubblica pone problemi di non lieve momento.** La riserva di giurisdizione opera, infatti, anche quale limite esterno all'attribuzione al potere giudiziario di poteri eccentrici o estranei al suo controllo che rischiano di coinvolgerlo in compiti securitari che non gli appartengono».

L'Italia ha di recente emanato il **d.lgs. 24/07/2023, n. 107** avente ad oggetto "Adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2021/784 del Parlamento europeo e del Consiglio, del 29 aprile 2021, relativo al contrasto della diffusione di contenuti terroristici on-line[243]. Ai sensi dell'art. 3 del d.lgs. 107/2023 l'autorità competente a emettere un ordine di rimozione nei confronti di un prestatore di servizi di hosting ai sensi dell'articolo 3, paragrafo 1, del regolamento, quando i **contenuti terroristici di cui all'articolo 2, punto 7) del regolamento sono riconducibili a un delitto con finalità di terrorismo**, è l'ufficio del **pubblico ministero** competente in base alle disposizioni del codice di procedura penale. Fuori dei casi di cui al primo periodo, l'ordine di rimozione è emesso dall'ufficio del pubblico ministero del tribunale del capoluogo del distretto che ha acquisito per primo la notizia relativa alla presenza sulle reti di telecomunicazioni disponibili al pubblico di contenuti terroristici. Il pubblico ministero informa immediatamente il Procuratore nazionale antimafia e antiterrorismo della ricezione della notizia di cui al comma 1 e ai fini della emissione dell'ordine di rimozione, il pubblico ministero acquisisce ogni necessario elemento informativo e valutativo, anche presso il C.A.S.A. Il pubblico ministero può, con decreto motivato, ritardare l'emissione dell'ordine di rimozione quando sia necessario per acquisire rilevanti elementi probatori ovvero per l'individuazione o la cattura dei responsabili dei delitti di cui al comma 1. L'ordine di rimozione è adottato con decreto motivato ed è portato a conoscenza dei destinatari preferibilmente per il tramite di agenti o ufficiali di polizia giudiziaria appartenenti all'organo del Ministero dell'interno per la sicurezza e la regolarità dei servizi di telecomunicazione. In caso di contenuti generati dagli utenti e ospitati su piattaforme riconducibili a soggetti terzi, è disposta la rimozione dei soli specifici contenuti illeciti. Prima di adottare i decreti indicati ai commi 5 e 6, il pubblico ministero informa il Procuratore nazionale antimafia e antiterrorismo. Ferma l'applicazione delle sanzioni di cui all'articolo 7, in caso di mancato adempimento, si dispone l'interdizione dell'accesso al dominio internet nelle forme e con le modalità di cui all'articolo 321 del codice di procedura penale, garantendo comunque, ove tecnicamente possibile, la fruizione dei contenuti estranei alle condotte illecite.

I **prestatori di servizi di hosting** che hanno ricevuto l'ordine di rimozione e i **fornitori dei contenuti** che, in conseguenza dell'ordine, sono stati rimossi o resi inaccessibili, nei dieci giorni successivi alla conoscenza del provvedimento, possono presentare opposizione innanzi al giudice per le indagini preliminari, che provvede con ordinanza in camera di consiglio a norma dell'articolo 127 del codice di procedura penale. Nondimeno, il ricorso per cassazione avverso l'ordinanza è ammesso unicamente per violazione di legge. L'art. 4 delinea la **procedura per gli ordini di rimozione transfrontalieri**, definendo la competenza del **GIP presso il tribunale del capoluogo del distretto** in cui il prestatore di servizi di hosting ha lo stabilimento principale o in cui il rappresentante legale del prestatore di servizi di hosting risiede o è stabilito. Il giudice dispone che copia dell'ordine di rimozione transfrontaliero sia trasmesso al **Procuratore nazionale antimafia e antiterrorismo** immediatamente e, comunque, prima di assumere le decisioni indicate al primo periodo. Le decisioni di cui all'articolo 4, paragrafi 3 e 4, del regolamento sono assunte, **sentito il pubblico ministero**, con decreto motivato. Nel caso previsto dall'articolo 4, paragrafo 4, del regolamento, avverso il decreto il prestatore di servizi di hosting e il fornitore di contenuti che hanno presentato la richiesta di esame dell'ordine di rimozione possono proporre ricorso per cassazione unicamente per violazione di legge. Il ricorso è proposto, a pena di decadenza, entro dieci giorni dal deposito del decreto. Il d.lgs. 107/2023 delinea un sistema articolato di **sanzioni amministrative e penali** per assicurare il rispetto di specifici obblighi delineati nel d.lgs. citato e nel Regolamento.

Tra essi anche l'**obbligo della conservazione dei contenuti terroristici rimossi o il cui accesso è stato disabilitato**, in linea con le disposizioni di cui all'articolo 6 del regolamento; i prestatori di servizi di hosting, infatti, devono conservare i contenuti terroristici rimossi o il cui accesso è stato disabilitato a seguito di un ordine di rimozione o di misure specifiche in conformità

[243] Per un primo commento cfr. A. CISTERNA, *Un modello di contrasto efficace anche per le altre minacce in rete*, in Guida al Diritto n, 38, p. 39-45.

dell'articolo 3 o 5, come pure i relativi dati rimossi in conseguenza della rimozione di tali contenuti terroristici quando siano necessari per: 1) i procedimenti di ricorso amministrativo o giurisdizionale, la gestione dei reclami ai sensi dell'articolo 10 contro una decisione di rimuovere o di disabilitare l'accesso ai contenuti terroristici e i relativi dati; 2) la prevenzione, l'accertamento, l'indagine o il perseguimento di reati di terrorismo. E 'sanzionato anche la trasgressione da parte dei *providers* degli **obblighi di trasparenza** di cui all'articolo 7 del Regolamento. I prestatori di servizi di *hosting* sono tenuti, in particolare, a definire chiaramente nelle loro **condizioni contrattuali** la loro politica volta a contrastare la diffusione di contenuti terroristici, che include, se del caso, **una valida spiegazione del funzionamento delle misure specifiche, compreso, ove applicabile, l'uso di strumenti automatizzati**. I prestatori sono tenuti a pubblicare annualmente **relazioni sulla trasparenza** che contengono almeno: informazioni sulle misure intraprese dal prestatore di servizi di *hosting* per quanto concerne l'individuazione e la rimozione o la disabilitazione dell'accesso a contenuti terroristici; informazioni sulle misure intraprese dal prestatore di servizi di *hosting* per contrastare la ricomparsa *online* di materiale che in precedenza sia stato rimosso o il cui accesso è stato disabilitato perché era stato considerato come integrante contenuti terroristici, in particolare ove siano stati utilizzati strumenti automatizzati; il numero di elementi integranti contenuti terroristici che sono stati rimossi o il cui accesso è stato disabilitato, a seguito di ordini di rimozione o misure specifiche e il numero di ordini di rimozione i cui contenuti non sono stati rimossi, o il cui accesso non è stato disabilitato, unitamente ai relativi motivi; il numero e l'esito dei reclami trattati dal prestatore di servizi di *hosting*; il numero e l'esito dei procedimenti di ricorso giurisdizionale o amministrativo avviati dal prestatore di servizi di *hosting*; il numero di casi in cui quest'ultimo è stato tenuto a ripristinare i contenuti, o l'accesso agli stessi, a seguito di procedimenti di ricorso giurisdizionale o amministrativo; il numero di casi in cui il prestatore di servizi di *hosting* ha ripristinato i contenuti, o l'accesso agli stessi, dopo aver esaminato un reclamo da parte del fornitore di contenuti.