Eidgenössische Finanzmarktaufsicht FINMA
Autorité fédérale de surveillance des marchés financiers FINMA
Autorità federale di vigilanza sui mercati finanziari FINMA
Swiss Financial Market Supervisory Authority FINMA

Comunicazione FINMA sulla vigilanza 08/2024

Governance e gestione del rischio nell'utilizzo dell'intelligenza artificiale

18 dicembre 2024

Laupenstrasse 27, 3003 Bern, Tel. +41 (0)31 327 91 00, www.finma.ch

Indice

1 Introduzione

L'impiego dell'intelligenza artificiale (IA) sul mercato finanziario è in aumento[1]. Per gli assoggettati alla vigilanza ciò offre opportunità, ma comporta anche rischi. Con la presente Comunicazione sulla vigilanza la FINMA richiama l'attenzione sui corrispondenti rischi, ma anche sulla necessità di identificarli, valutarli, gestirli e monitorarli in maniera adeguata.

Attualmente in Svizzera non esiste una legislazione specifica in materia di IA. Nel diritto dei mercati finanziari, i requisiti prudenziali improntati alla neutralità tecnologica e basati sui principi per una *governance* efficace e una gestione efficace dei rischi comprendono i rischi derivanti dall'utilizzo dell'IA. Come richiesto anche in ambito internazionale, la FINMA si attende che gli assoggettati alla vigilanza che utilizzano l'IA si confrontino attivamente con le relative ripercussioni sul loro profilo di rischio e che adeguino di conseguenza l'orientamento della loro *governance*, della loro gestione del rischio e dei loro sistemi di controllo. Al riguardo occorre considerare, oltre alle dimensioni, alla complessità, alla struttura e al profilo di rischio degli assoggettati, in particolare l'essenzialità delle applicazioni di IA utilizzate e la probabilità che i rischi derivanti dall'utilizzo di tali applicazioni si verifichino[2].

2 Constatazioni effettuate nel quadro della vigilanza

I rischi derivanti dall'utilizzo dell'IA risiedono principalmente nell'ambito dei rischi operativi[3], in particolare nei rischi di modello (p.es. robustezza carente, correttezza, distorsione e spiegabilità) come pure nei rischi informatici e nei cyber-rischi. Risultano anche da una crescente dipendenza di terzi, per esempio fornitori di soluzioni hardware, modelli o servizi cloud in un mercato sempre più concentrato[4]. Infine, sussistono rischi giuridici di reputazione come pure difficoltà nell'attribuzione delle responsabilità a causa delle azioni

[1] Per quanto concerne la diffusione dell'intelligenza artificiale sul mercato finanziario, cfr. «FSB, The Financial Stability Implications of Artificial Intelligence», 14.11.2024 (di seguito: FSB), pag. 3 segg.

[2] Possibili fattori che influiscono sull'essenzialità di un'applicazione sono (l'elenco non è esaustivo): importanza per l'osservanza della legislazione in materia di mercati finanziari, ripercussioni finanziarie sull'impresa, rischi giuridici e di reputazione, rilevanza del prodotto per l'impresa, numero di clienti e investitori interessati, tipologie di clienti e investitori (retail/istituzionali), importanza del prodotto per i clienti e gli investitori, conseguenze in caso di errori o di interruzione del funzionamento. Possibili fattori che influiscono sulla possibilità che gli eventi associati ai rischi si verifichino sono (l'elenco non è esaustivo): complessità (p. es. spiegabilità, prevedibilità), tipologia e quantità dei dati utilizzati (p. es. dati non strutturati, integrità, pertinenza, dati personali), processi di sviluppo e di monitoraggio inadeguati, grado di autonomia e integrazione nei processi, dinamica (p. es. brevi cicli di calibrazione), collegamento in rete di più modelli, potenziale di attacchi o di interruzione del funzionamento (p. es. in aumento a causa dell'esternalizzazione).

[3] Cfr. art. 89 OFoP: è definito rischio operativo il pericolo di perdite consecutive all'inadeguatezza o all'inefficacia delle procedure interne, delle persone o dei sistemi oppure dovute a eventi esterni.

[4] Cfr. anche FSB, pag. 16 segg.

autonome e difficili da spiegare di questi sistemi e della dispersione delle responsabilità per le applicazioni di IA tra gli assoggettati alla vigilanza.

Di seguito sono indicate per esempio misure per la gestione specifica dei rischi derivanti dalle applicazioni di IA che la FINMA ha constatato nel quadro della vigilanza continua, segnatamente nei colloqui di vigilanza e nei primi controlli in loco specifici. L'obiettivo è quello di supportare gli assoggettati alla vigilanza nell'individuazione, nella valutazione, nella gestione e nel monitoraggio dei rischi derivanti dalle applicazioni di IA interne ed esterne.

2.1 *Governance*

La FINMA ha constatato che gli assoggettati alla vigilanza si concentrano innanzitutto sui rischi legati alla protezione dei dati e in misura minore sui rischi di modello come robustezza carente e correttezza, distorsioni *(bias)*, stabilità carente e spiegabilità. Inoltre, lo sviluppo delle applicazioni di IA avviene spesso in maniera decentralizzata, pertanto risulta difficile applicare standard coerenti, assegnare le responsabilità in modo chiaro e a dipendenti con competenze ed esperienza adeguate e gestire tutti i rischi rilevanti. Per le applicazioni e i servizi acquistati esternamente, gli assoggettati alla vigilanza hanno talvolta avuto difficoltà a determinare se l'IA fosse inclusa, quali dati e metodi venissero impiegati e se esistesse una *due diligence* sufficiente.

La FINMA ha valutato se presso gli assoggettati con molte applicazioni o con applicazioni essenziali fosse stata predisposta una *governance* in materia di IA, che comprenda fra le altre cose un inventario gestito a livello centrale, incluse una classificazione del rischio e le misure che ne conseguono, la determinazione delle competenze e delle responsabilità nello sviluppo, nell'implementazione, nel monitoraggio e nell'utilizzo dell'IA, disposizioni sui test dei modelli e sui controlli di supporto del sistema, standard di documentazione e ampie misure di formazione. In caso di esternalizzazione ha valutato se gli assoggettati avessero implementato ulteriori test, controlli e clausole contrattuali che disciplinano le responsabilità e le questioni relative alla responsabilità e se si fossero assicurati che i terzi incaricati dispongano delle competenze e delle esperienze necessarie.

2.2 Inventario e classificazione dei rischi

La FINMA ha constatato che talvolta gli assoggettati alla vigilanza definivano l'IA in modo restrittivo per concentrarsi sui rischi presumibilmente grandi o nuovi. Per molti assoggettati alla vigilanza è risultato complicato garantire la completezza degli inventari, in quanto lo sviluppo e l'applicazione dell'IA sono spesso ampiamente diffusi all'interno dell'azienda e, da quando sono state introdotte le applicazioni di IA generativa, sono accessibili su larga scala. Inoltre, non tutti gli assoggettati alla vigilanza avevano definito criteri

coerenti per identificare le applicazioni che richiedevano particolare attenzione nella gestione dei rischi a causa della loro essenzialità, dei rischi specifici e della probabilità di verificarsi[5].

La FINMA ha valutato se presso gli assoggettati alla vigilanza era disponibile una definizione sufficientemente ampia di IA[6], in quanto anche le applicazioni classiche possono presentare rischi analoghi e i rischi analoghi devono essere gestiti allo stesso modo[7]. Ha pertanto valutato la presenza e la completezza degli inventari di IA come pure la classificazione del rischio delle applicazioni di IA.

2.3 Qualità dei dati

La FINMA ha constatato che in alcuni casi gli assoggettati alla vigilanza non avevano definito disposizioni né predisposto controlli per garantire la qualità dei dati nelle applicazioni di IA.

Spesso le applicazioni di IA apprendono in maniera automatizzata e senza l'intervento umano. La qualità dei dati è dunque in molti casi più importante rispetto alla scelta del modello concreto. Nel contempo, i dati possono essere errati, incoerenti, incompleti, non rappresentativi o obsoleti, e pertanto di cattiva qualità. I dati storici possono contenere distorsioni che si ripercuotono sulle previsioni future oppure possono non essere più rappresentativi della previsione a causa di un contesto modificato. Nel caso di soluzioni acquistate, spesso gli assoggettati alla vigilanza non possono influenzare in alcun modo i dati sottostanti o non li conoscono. Ciò può fare sì che non siano più adeguate per gli assoggettati o per la richiesta specifica e aumenta il rischio di un uso inconsapevole di dati deliberatamente manipolati. Da quando l'utilizzo dell'IA è aumentato, vengono inoltre analizzati più dati non strutturati, come testi e immagini, per i quali può essere più difficile valutare la qualità.

La FINMA ha valutato se, nelle loro direttive e linee guide interne, gli assoggettati alla vigilanza hanno definito disposizioni per garantire la completezza, la correttezza e l'integrità dei dati nonché la disponibilità e l'accesso ai dati.

[5] Tendenzialmente i rischi sono elevati se l'IA viene impiegata per l'osservanza del diritto prudenziale o per l'esecuzione di funzioni critiche oppure se la clientela o i collaboratori sono fortemente colpiti dai relativi risultati. I criteri per la classificazione dovrebbero essere stabiliti dagli assoggettati alla vigilanza.

[6] Cfr. approccio alla definizione dell'OECD: OECD, «Explanatory Memorandum on the Updated OECD Definition of an AI System», OECD Artificial Intelligence Papers, March 2024 (No. 8).

[7] Di per sé l'IA non è un'applicazione che presenta un rischio elevato. Il rischio ad essa associato dipende dalla complessità, dall'adattabilità e dall'autonomia della corrispondente applicazione, del suo ambito di applicazione e della sua integrazione nei processi.

2.4 Test e monitoraggio costante

La FINMA ha in alcuni casi constatato presso gli assoggettati carenze nella scelta degli indicatori di performance, nei test e nel monitoraggio costante.

La FINMA ha valutato se gli assoggettati alla vigilanza hanno messo a punto test per garantire la qualità dei dati e il buon funzionamento delle applicazioni di IA, che comprendono una verifica della precisione, della robustezza, della stabilità ed eventualmente delle distorsioni[8]. Ha inoltre valutato se gli specialisti nel rispettivo ambito di applicazione hanno formulato domande e aspettative prestabilite e se sono stati impostati indicatori predefiniti della performance per valutare la capacità dell'applicazione di IA di raggiungere gli obiettivi prefissati[9]. Per quanto concerne i controlli periodici, la FINMA ha valutato per esempio se gli assoggettati alla vigilanza avevano definito valori soglia o altri metodi di validazione per garantire la correttezza e la qualità continua dei risultati[10]. Ha altresì esaminato se gli assoggettati alla vigilanza monitorano le modifiche dei dati di input, per garantire l'applicabilità dei modelli anche in un contesto modificato (riconoscimento e trattamento della deriva dei dati). Il monitoraggio include anche l'analisi dei casi in cui l'*output* è stato ignorato o modificato dagli utenti, in quanto tali correzioni manuali possono rivelare vulnerabilità. Infine, la FINMA ha valutato se gli assoggettati alla vigilanza stiano già formulando riflessioni sull'individuazione e sulla gestione delle eccezioni.

2.5 Documentazione

La FINMA ha constatato che, in alcuni casi, gli assoggettati non hanno elaborato disposizioni a livello centrale in materia di documentazione e che l'attuale documentazione a volte non è sufficientemente dettagliata e orientata al destinatario.

Per le applicazioni essenziali la FINMA ha valutato se nella documentazione gli assoggettati alla vigilanza trattano lo scopo dell'applicazione, la selezione e la preparazione dei dati, la scelta del modello, la misurazione della performance, le ipotesi, le limitazioni, i test e i controlli come pure le soluzioni di *fallback*. Nella selezione dei dati la FINMA ha considerato se gli assoggettati

[8] Esistono molteplici test per valutare la performance e i risultati di un'applicazione. Ciò comprende fra le altre cose i test in cui gli utenti conoscono il risultato corretto e verificano se l'applicazione lo fornisce (ad es. *backtesting, out-of-sample testing*), test concepiti per capire come si comporta l'applicazione in determinati casi limite (ad es. analisi di sensitività o *stress test*), test con dati di input errati (ad es. *adversarial testing*) o test rispetto a ulteriori modelli di *benchmark* più semplici. Inoltre, i test possono essere utilizzati per valutare i potenziali limiti dell'applicazione e per verificare la "ripetibilità" dei risultati.

[9] Più l'applicazione è essenziale e complessa e meno si conosce del funzionamento del sistema o dei dati sottostanti, più è importante valutare in maniera continua, prima dell'utilizzo produttivo, se l'applicazione funziona conformemente al suo scopo in caso di modifiche, soprattutto a causa dell'adattabilità delle applicazioni odierne. È inoltre importante formulare riflessioni sui meccanismi di *fallback*, in modo tale da essere preparati qualora l'IA si sviluppi in una direzione indesiderata e non soddisfi più gli obiettivi originariamente definiti.

[10] A tale scopo possono risultare utili campioni, *backtesting*, test predefiniti o *benchmarking*.

alla vigilanza presentavano le fonti dei dati e i controlli della qualità dei dati, comprese l'integrità, la correttezza, la pertinenza, la rilevanza, le distorsioni e la stabilità. Inoltre ha esaminato in che modo gli assoggettati alla vigilanza garantiscono la robustezza e l'affidabilità come pure la tracciabilità dell'applicazione e se effettuano adeguatamente la classificazione in una categoria di rischio come pure la relativa motivazione e verifica.

2.6 Spiegabilità

La FINMA ha constatato che spesso i risultati non possono essere compresi, spiegati o riprodotti e pertanto non possono essere valutati criticamente.

Se hanno dovuto essere motivate decisioni nei confronti degli investitori, dei clienti, dei collaboratori, della vigilanza o della società di audit, la FINMA ha valutato in modo approfondito la spiegabilità delle applicazioni. Ciò include, tra le altre cose, la comprensione dei driver delle applicazioni o il comportamento in diverse condizioni per poter valutare la plausibilità e l'attendibilità dei risultati.

2.7 Verifica indipendente

Non sempre la FINMA ha constatato una chiara delimitazione tra lo sviluppo di applicazioni di IA e la verifica indipendente.

Inoltre, ha osservato che solo pochi assoggettati alla vigilanza svolgono una verifica indipendente di tutto il processo di sviluppo del modello da parte di personale qualificato a tale scopo, per identificare in maniera coerente i rischi di modello e ridurli.

In caso di applicazioni essenziali, la FINMA ha valutato se la valutazione indipendente comprende la formulazione di un giudizio obiettivo, esperto e imparziale sull'adeguatezza e l'affidabilità di una procedura per un determinato caso d'uso e se i risultati della verifica indipendente sono stati considerati nello sviluppo dell'applicazione.

3 Prospettive

La comprensione del rischio nell'utilizzo dell'IA da parte degli assoggettati alla vigilanza è in fase di sviluppo. Sulla base delle esperienze maturate nella vigilanza e in linea con gli sviluppi internazionali rilevanti, anche la FINMA svilupperà ulteriormente le proprie aspettative nei confronti di una *governance* e una gestione del rischio adeguate da parte degli assoggettati in relazione con l'IA e, se necessario, creerà trasparenza sul mercato. Come per altri fattori di rischio rilevanti, la FINMA intende adottare un approccio

improntato alla neutralità tecnologica, proporzionale e intersettoriale, tenendo conto delle principali differenze tra i settori e degli standard internazionali.