



RAISE

Reportistica Antiriciclaggio Integrata e Sintesi Estrattiva

Integrated Anti-Money Laundering Reporting and Extractive Summarization

ECONDAT Knowledge Session
2 October 2025

Leveraging Retrieval-Augmented Generation for AML/CFT Compliance: The RAISE System

Daniele Licari

IT Development Directorate, Bank of Italy

daniele.licari@bancaditalia.it

Key Points of the Presentation

- Background
- What is RAISE?
- Methodology
- Results
- Conclusions

The logo for RAISE, featuring the word "RAISE" in a bold, dark blue, sans-serif font. A small green square is positioned between the 'A' and 'I'.

Reportistica Antiriciclaggio Integrata e Sintesi Estrattiva

Background:

AML Supervision and Regulation Unit

Supervisory Model

- **Risk-based approach**, guided by annual assessments
- Relies on **AML/CFT reports** and **Risk Assessments** sent by intermediaries

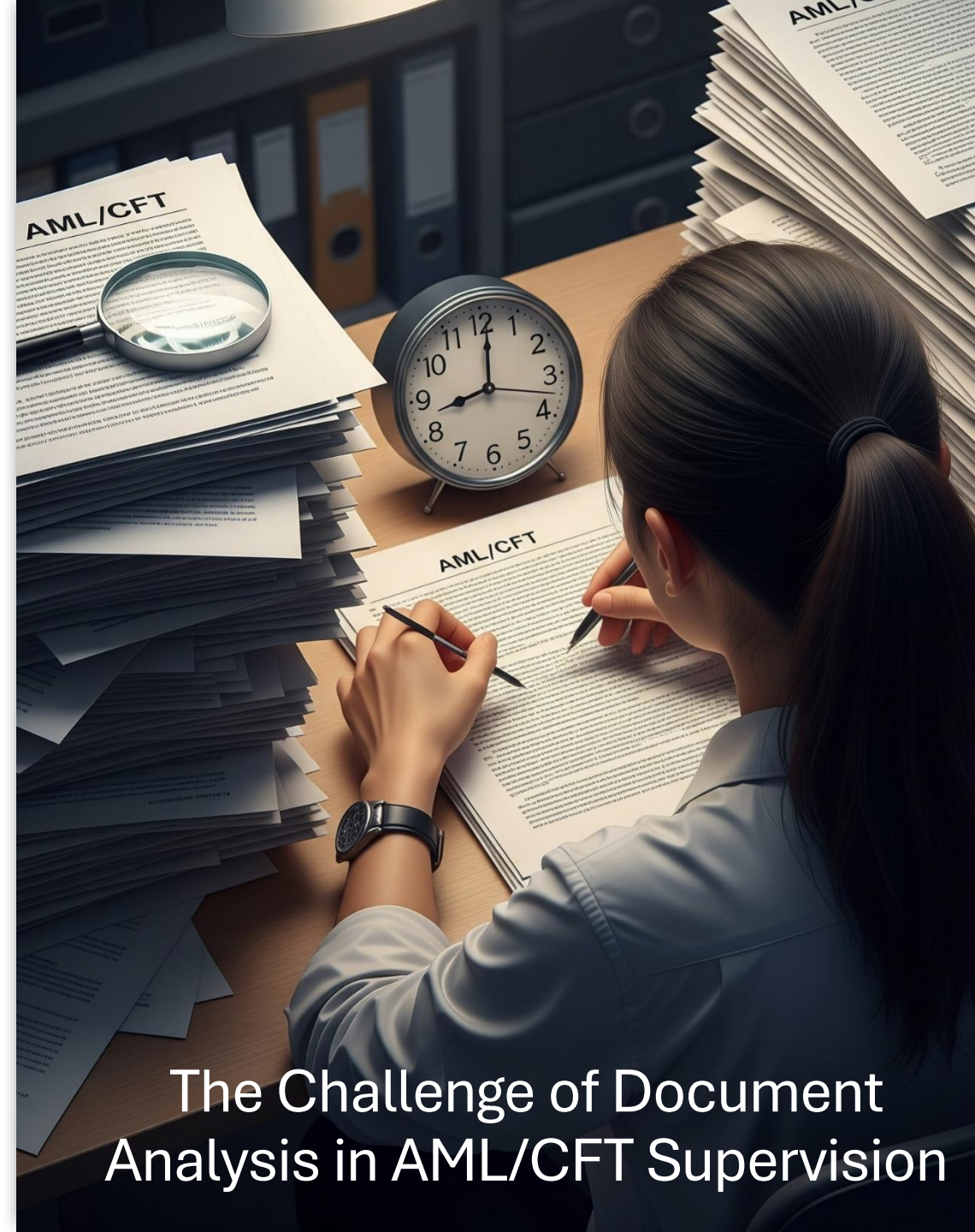
Current Challenges

- Submissions are **unstructured** and vary in detail
- Reviews are **manual, complex, and resource-intensive**
- **>3 person-days per intermediary** just for extraction & summarization*

Project scope

- 120 documents from 87 intermediaries; test set: 20 intermediaries balanced across sectors

* based on internal analyses conducted on the reports and self-assessments from various categories of supervised intermediaries.



The Challenge of Document
Analysis in AML/CFT Supervision

What is RAISE?

RAISE (Reportistica Antiriciclaggio Integrata e Sintesi Estrattiva) is an AI-based system that supports the automatic analysis of AML/CFT reports, **extracting relevant information from documents and generating structured summaries for each intermediary.**

Objectives

- **Accelerate** the analysis of documents during the annual review cycle
- **Support** analysts in producing draft summary reports
- **Increase** consistency of intermediary profiles
- **Create** structured datasets from unstructured texts



Attività Prevalente

100% 100.000 clienti

Operazioni Online

100%

100% 100.000 clienti

Promotori Finanziari

55

one Aziendale

figura e gli strumenti comuni. Le attività del Gruppo che operano con un

za con le linee guida di Gruppo e sotto la supervisione funzionale della



one dei Rischi

Incremento Operazioni Dispositive

15%

Vulnerabilità del Sistema di Controllo

3 su 4

Relazione e Verifiche

Le principali criticità emerse nel processo di adeguata verifica includono l'attuale sistema informatico di monitoraggio delle transazioni, che opera su regole statiche e generiche, risultando in un elevato numero di falsi positivi. Questo ha comportato un dispendio di risorse per l'analisi manuale di segnalazioni irrilevanti. Nel 2023, sono stati generati circa 17.000 alert dal sistema Giove, dei quali oltre il 95% classificati come falsi positivi dopo la alcune procedure AML, tra Capogruppo e controllate, in particolare nella procedura di adeguata verifica e profilatura del rischio cliente. Questo ha portato a un approccio non omogeneo nella classificazione del rischio e a monitoraggio continuo dei clienti ad alto rischio la cosiddetta vulnerabilità, con documentazione incompleta o non aggiornata per alcuni clienti PEP e conti di clientela privata non condotti con operatività complessa. È di automare una di routine del sistema informatico. Infine, abbiamo i tassi di partecipazione alla formazione AML, siano essi, l'autovalutazione ha evidenziato aree dove la conoscenza del personale potrebbe essere rafforzata gestore di contante complesso.

A seguito delle verifiche, sono state intraprese diverse azioni correttive, tra cui l'implementazione del progetto "Nuovo Monitoraggio" (Giove2), la revisione delle policy di gruppo, la sostituzione dei fornitori PEP, la formazione per i clienti ad alto rischio e l'aggiornamento di un KPI di completezza dossier ai gestori privati. Durante il 2023, la Banca ha tenuto costantemente informata Banca d'Italia sui progressi del piano di rimedio, in finale in novembre. La Vigilanza ha espresso valutazione positiva preliminare sugli stessi compiti, pur attendendo di verificare gli effetti nel tempo. Non sono state ricevute sanzioni amministrative né contestazioni per



Governance e Procedure

La metodologia di pianificazione strategica prevede un'analisi annuale dei fabbisogni formativi basata su dati dell'autovalutazione, evoluzioni normative, introduzione di nuovi prodotti e processi, risultati dei controlli interni. Il Piano Formativo Antiriciclaggio definisce interventi formativi per target di destinatari e per modalità didattiche. Nel 2023, è stata posta enfasi nella formazione specialistica in tema di adeguata verifica rafforzata e nella APC ha garantito di coinvolgere il personale professionale e di sviluppare un programma interno di AML, Talent Development in collaborazione con HR.

L'adeguata verifica della clientela è il primo pilastro nella prevenzione del riciclaggio, comprendendo identificazione, verifica dell'identità, acquisizione di informazioni sullo scopo del rapporto e sulla provenienza dei fondi. Nel 2023, la Banca XYZ ha effettuato l'adeguata verifica su tutti i nuovi clienti seguiti e ha provveduto con la revisione periodica della clientela esistente secondo gli intervalli stabiliti. Per ciascun nuovo cliente è stata eseguita un'analisi di rischio basata su dati di rischio e di 4% tramite outcoming di rischio con strumenti digitali. Non sono state riscontrate situazioni di mancato adeguata verifica. L'Archivio Unico Informatico ha registrato correttamente la persona fisica, la Banca ha identificato e verificato l'identità dei titolari effettivi secondo quanto previsto dalla normativa. A ogni nuovo cliente è stato assegnato un profilo di rischio iniziale tramite un modello informatico. L'EU% è stato classificato a rischio basso, il 15% Medio e il 3% Alto. La classificazione Alto è scattata in particolare per tutte le persone identificate come PEP o provenienti da paesi con carenze AML strategiche. Per quanto parte di un funzionario della Funzione AML, prima di consentire piena operatività.

La conservazione minima di 10 anni è garantita da procedure automatiche. La revisione periodica della clientela esistente è assicurata per clienti alto rischio, trimestrale per medio e quinquennale per basso rischio. Le categorie attività di riciclaggio o finanziamento del terrorismo includono Persone Politicamente Esposte, clienti non residenti, società fiduciarie o trust, soggetti operanti in settori sensibili e strutture e strutture complesse.

La rilevazione delle operazioni anomale e la successiva eventuale Segnalazione di Operazioni Sospette rappresentano il cuore dell'attività antiriciclaggio quotidiana. Nel 2023, la Banca XYZ ha rafforzato i propri processi sistemi di monitoraggio basati su scenari predefiniti ha generato circa 17.000 alert automatici, e nel 2023 sono state segnalate segnalazioni interne non informative. Il team dedicato ha effettuato un'analisi approfondita di ogni alert verifiche. Le tipologie di anomalie riscontrate hanno riguardato operazioni con l'elenco, utilizzo improprio del contante e movimentazioni non coerenti col profilo. Nel corso del 2023, il Gruppo XYZ ha trascorso cinque milioni di euro per la segnalazione di queste operazioni includono posizioni con saldo conto modesto che ricevono grandi bonifici, sospetto di furto di fondi su finanziamenti leasing e strutture di layering con

La Banca ha valutato e implementato tecnologie emergenti per migliorare l'efficienza dei processi AML, come modelli di machine learning e modelli di network analysis. Inoltre, sono stati valutati servizi di blockchain analitica e di alcune richieste della Vigilanza, la Banca ha definito un Piano di Azione Coesistente approvato dal Consiglio di Amministrazione. Nel corso del 2023, la Funzione AML ha avviato e in parte completato le azioni policy di gruppo, effettuando la consultazione dei fornitori PEP e facendo formazione specialistica. Sono state introdotte una checklist rafforzata di outcoming per i clienti ad alto rischio e un KPI di completezza dossier a



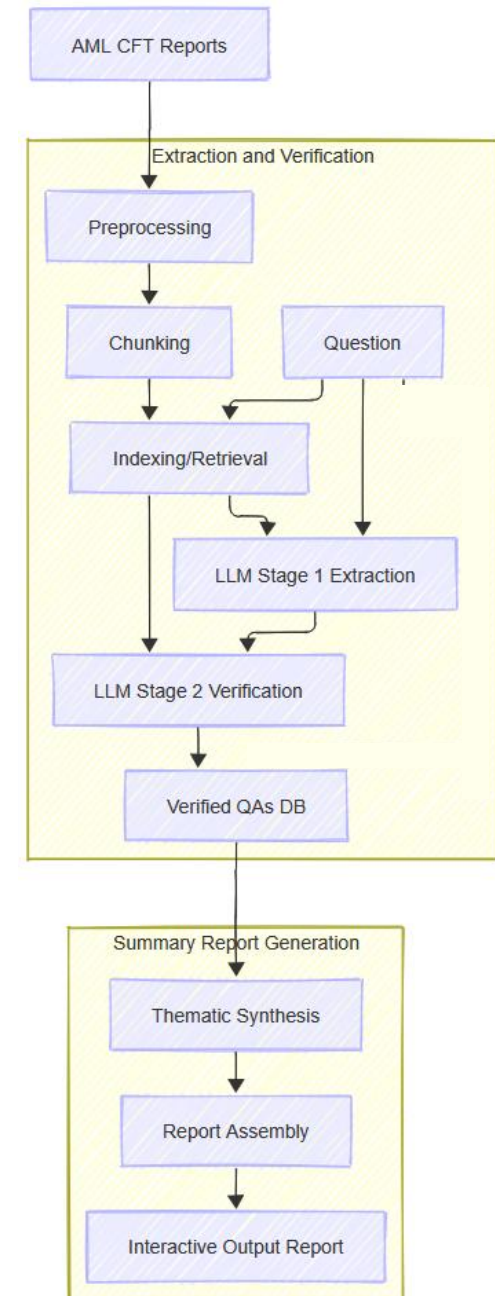
Methodology: Phases Overview

1. Extraction & Verification

- Preprocess documents → chunk text → retrieve relevant passages → LLM extracts answers → verification step → **verified Q&As**

2. Summary Report Generation

- Synthesize verified Q&As into structured sections: Activity, Organization, Risk, Controls, Governance
- Produce **interactive HTML** report with expandable answers and full source traceability



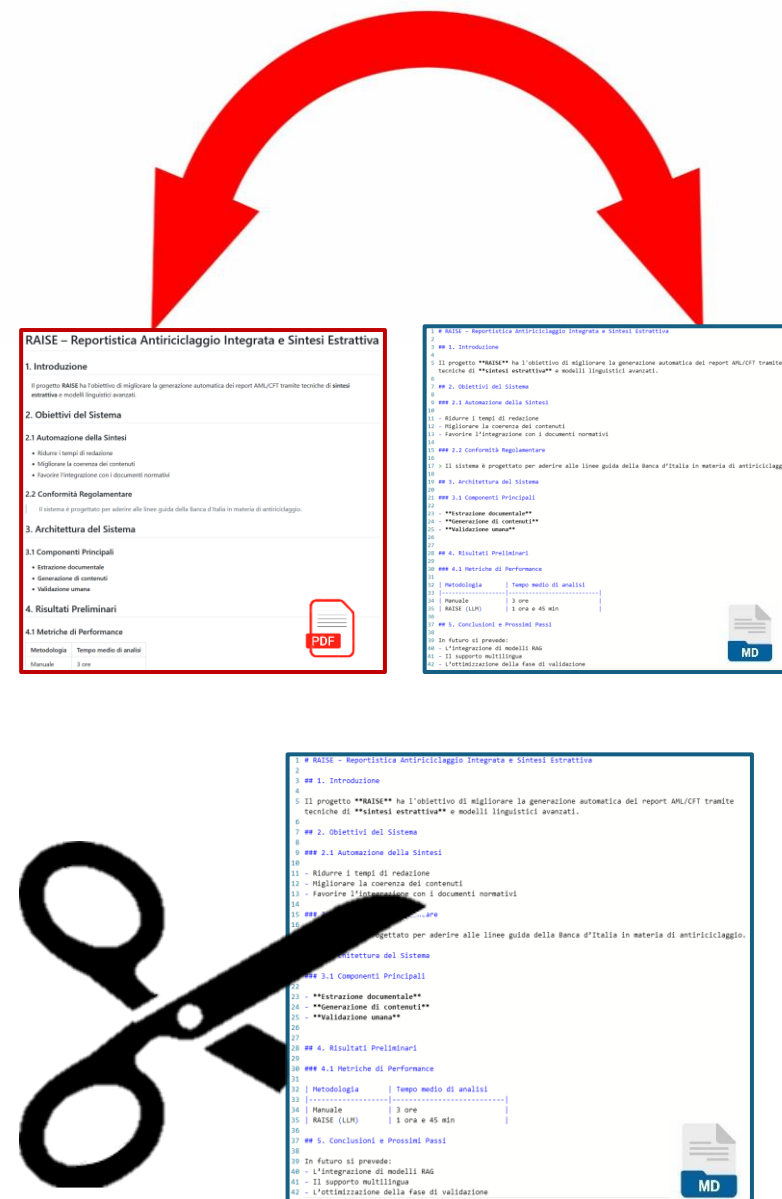
Data Preparation & Segmentation

Document prep: PDF → Markdown, preserving titles, lists, tables

Hierarchical segmentation:

- **Block segmentation:** divide text by titles & paragraphs, preserving context
- **Sentence segmentation:** create smaller, cohesive units within chunk size constraint

Benefit: LLM can efficiently locate relevant “pieces” of the document, improving accuracy and extraction efficiency



Targeted Questions & Annotation

58 predefined questions across 5 AML/CFT areas: Activity, Organization, Risk, Controls, Governance. Designed by domain experts → ensures **comprehensive, consistent coverage**

Annotation:

- 28 analysts annotated 120 documents → 6,524 annotations (~3,000 Q&A pairs)
- Iterative workflow for **consistency & quality control**
- Representative examples selected for **few-shot learning**, optimizing LLM performance

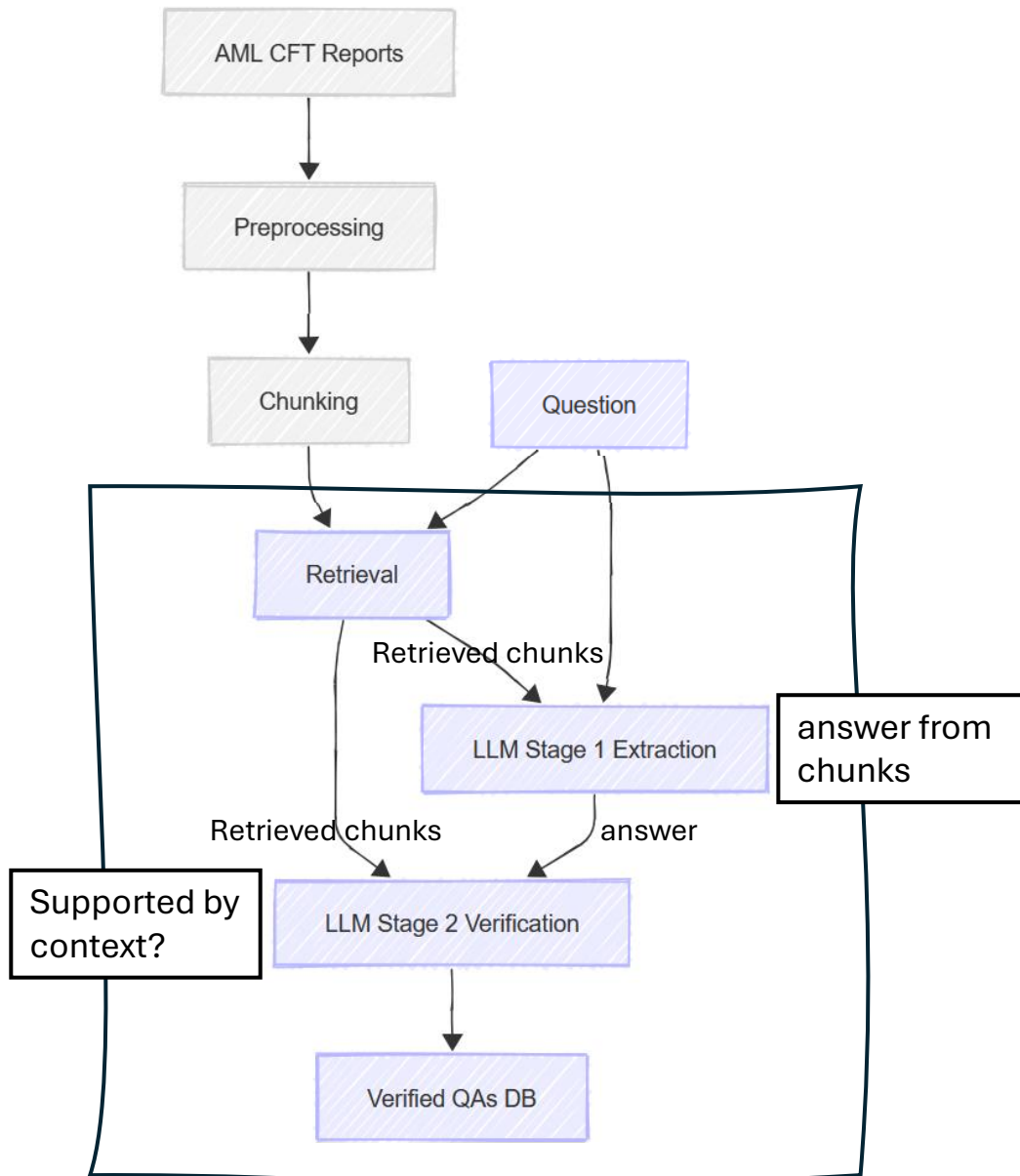
Area	N. Questions
Main Business Activity Framing of the intermediary and business model.	7
Corporate Organization Structure of the AML function and the roles of corporate bodies	10
Risk Self-Assessment Internal process for assessing ML/TF (Money Laundering/Terrorist Financing) risk.	15
Reporting and Controls Activities carried out by the AML function and the controls performed.	9
Governance and Procedures AML management at the level of governance, processes, technological developments, and interactions with authorities.	17
Tot	58

1	A2	A3	A4	A5	A6	A7	A8	B1	B2	B3	B4	B5	B6	B7	B8
11	B12	B13	B14	B15	B16	B17	B18	B19	C1	C2	C3	C4	C5	C6	C7
18	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25
13	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18
	D19	D20	D21	D22											

1. PREMESSA. OBIETTIVO E CONTESTO NORMATIVO DI RIFERIMENTO 7
2. METODOLOGIA 8
3. FASI DEL PROCESSO DI VALUTAZIONE E ATTORI COINVOLTI 8
3.1 Fase 1- Raccolta dati e informazioni 8
3.2 Fase 2- Elaborazione 9
3.3 Fase 3- Predispersione della Relazione annuale con allegata Autovalutazione 10
3.4 Fase 4- Discussione, condivisione e approvazione 10
3.5 Fase 5- Verifica stato di attuazione precedenti iniziative 10
4. RISULTATI DELL'ANALISI 11
4.1 Determinazione del Rischio inerente 11
4.1.1 Individuazione dell'area di business 11
4.1.2 Individuazione dei fattori di rischio 11
4.1.3 Validazione del rischio inerente 15
4.2 Analisi di vulnerabilità dei presidi 15
4.3 Rischio residuo 17
5. STATO AVANZAMENTO AZIONI DI RIMEDIO INDICATE NELL'ANNO PRECEDENTE 19
6. AZIONI DI RIMEDIO 21
8. ALLEGATI 22

 **Label Studio**
on-prem open-source
annotation tool

autorizzata all'esercizio dell'attività disciplinata dalla L. 23/11/1939 n. 1906 con Decreto Ministeriale del 7 dicembre 2012 Iscritta nella sez.



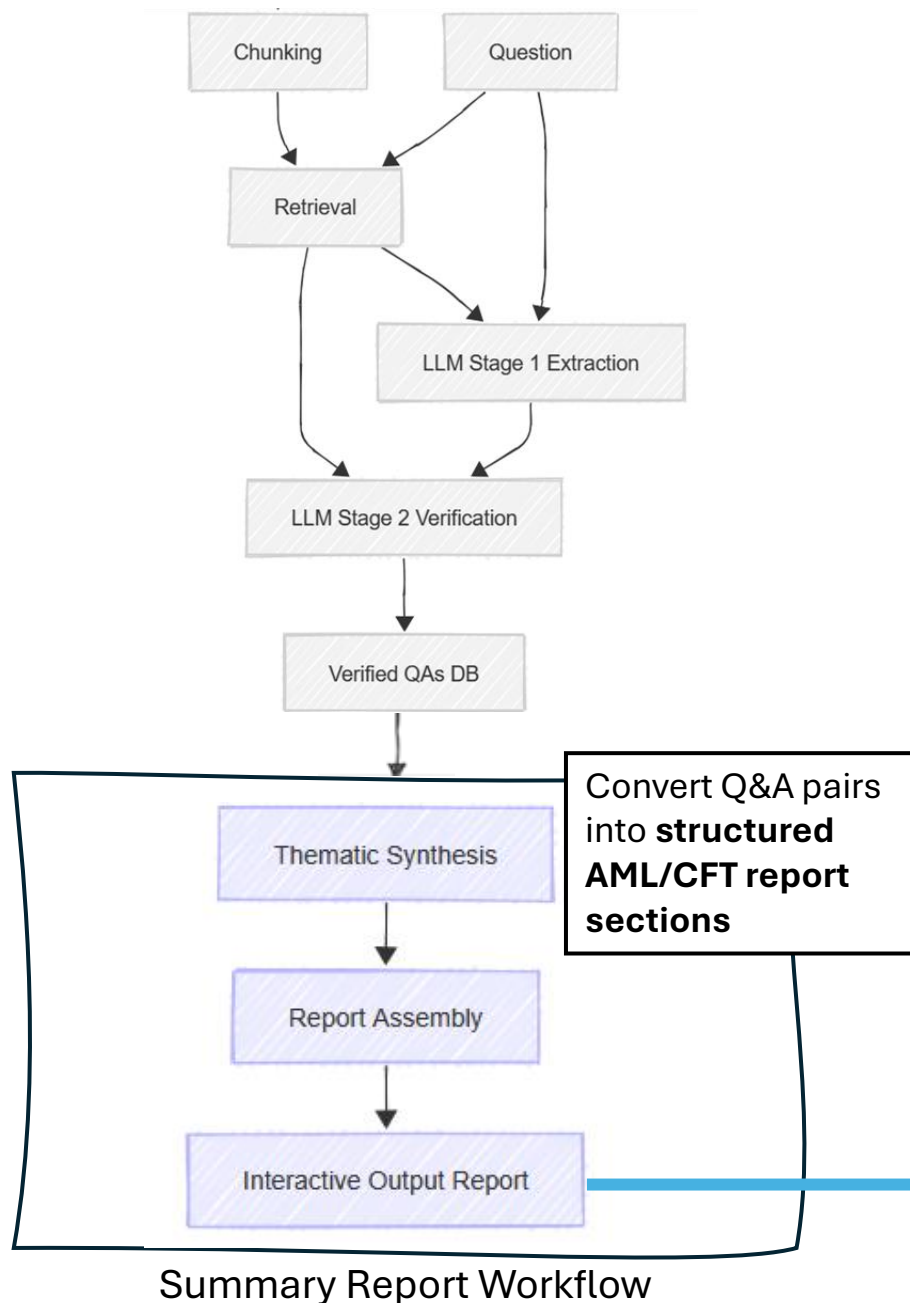
RAG & Verification Workflow

Extraction and Verification

- **Hybrid retrieval***: BM25 + domain-tuned BGE-m embeddings (Reciprocal Rank Fusion)
- **LLM extraction**: Mistral Small 3.1**, guided by few-shot examples, extracts answers, links to sources → transparency & traceability
- **Verification**: LLM checks each answer is fully supported by retrieved context
- **Segmentation & configuration**: stateless, chunk size 1,024 tokens, 5 chunks per question
- **Outcome**: Verified Q&A set per intermediary, ready for structured report generation

*BM25 + domain-tuned BGE-m3 embeddings, (4-epoch AdamW fine-tuning with contrastive loss on 2,100 Q&A pairs, validated on 300 pairs) combined with Reciprocal Rank Fusion

** Mistral-Small-3.1-24B-Instruct-2503



Summary Report Generation

- **Input:** Verified Q&A pairs, report topics
- **Process:** Synthesize into structured sections (Activity, Organization, Risk, Controls, Governance); generate **interactive HTML report**
- **Interactive Sections:** Expandable Q&A pairs per section

Attività Prevalente:

Il Gruppo Bancario XYZ è un istituto bancario commerciale attivo principalmente nel retail e corporate banking, con servizi di private banking dei rapporti con i clienti. La rete distributiva conta 120 filiali sul territorio nazionale, con una concentrazione in Lombardia, Veneto, Emilia-Romagna. I canali digitali riducono il contatto diretto e possono essere sfruttati per tentare aperture di rapporti con identità false. Nel 2023 il peso dei canali a distanza è aumentato.

Extraction Process

Organizzazione Aziendale:

La Funzione Anti Financial Crime (AFC) del Gruppo XYZ, che presidia i rischi di AML/CFT, è collocata all'interno della Direzione Compliance. La Funzione "Operazioni Sospette" per l'intero Gruppo XYZ, Rossi riporta gerarchicamente al Chief Compliance Officer del Gruppo, ma riferisce funzionale al Chief Risk Officer. La Funzione "Operazioni Sospette" rimane oggetto di monitoraggio e possibile ulteriore upgrading in caso di espansioni del business o inasprimenti normativi. Nel corso del 2023, la Funzione "Operazioni Sospette" ha implementato un nuovo sistema di monitoraggio e nell'analisi avanzata dei dati. Il modello di gestione dei presidi AML adottato dal Gruppo XYZ, che prevede la centralizzazione dei presidi propri, hanno realizzato le attività in coerenza con le linee guida di Gruppo e sotto la supervisione funzionale della Funzione AML centralizzata.

Extraction Process

Autovalutazione dei Rischi:

Il rischio inerente del Gruppo è determinato da diversi fattori, tra cui la tipologia delle operazioni, la complessità delle transazioni, la presenza di fenomeni di riciclaggio. Nel 2023, si è osservato un aumento del 2% del numero di operazioni sospette, che ha portato a un aumento del 10% del rischio inerente poiché riduce il conteggio delle operazioni transfrontaliere della Banca ha coinvolto paesi extra-UE classificati ad alto rischio. I principali fattori di rischio inerente identificati dall'intermediario sono i canali digitali (3 su 4). Il calcolo del rischio inerente combina indicatori quantitativi e qualitativi. Le maggiori criticità emerse includono vulnerabilità nei sistemi IT e nel monitoraggio delle operazioni sospette. Nel corso del 2023, la Funzione AML ha completato azioni correttive, tra cui l'implementazione del progetto "Nuovo Monitoraggio Operazioni Sospette" e la revisione straordinaria di tutti i dossier PEP e High Risk pendenti. Nel corso del 2023, la Funzione AML ha avviato e in parte completato le azioni di rafforzamento del sistema di monitoraggio e consolidare la cultura aziendale di compliance.

Extraction Process

Relazione e Verifiche:

La funzione antiriciclaggio della Capogruppo effettua verifiche annuali in presenza dei clienti e monitoraggio degli investimenti per antiriciclaggio, e sul rispetto delle normative. Nel corso del 2023, la Banca ha rifiutato l'instaurazione del rapporto continuativo con i clienti che non hanno superato la classificazione del profilo di rischio di circa 300 clienti, che sono stati alzati da Medio a Alto rischio. L'attuale sistema informatico di monitoraggio delle transazioni, software "Giano", non è in grado di gestire la complessità delle operazioni sospette e la conformità nella declinazione di alcune procedure AML tra Capogruppo e controllati. L'autovalutazione ha messo in luce che il meccanismo di review periodica dei clienti è insufficiente e in parte un'insufficiente consapevolezza operativa su tale obbligo. Nel corso del 2023, il Gruppo XYZ ha trasmesso complessivamente 342 Segnalazioni di Sospette Operazioni (SSO) alle autorità competenti, fornendo tutto il materiale richiesto. Non vi sono stati annullamenti o contestazioni. Nel corso del 2023, la Funzione AML ha avviato e in parte completato le azioni di rafforzamento del sistema di monitoraggio e consolidare la cultura aziendale di compliance.

Extraction Process

Governance e Procedure:

Il management di alto livello, inclusi l'Amministratore Delegato e il Collegio Sindacale, è informato regolarmente sull'andamento dei presidi AML e sulle azioni correttive in corso e si promuove il coordinamento tra funzioni, come la condivisione di casistiche di operazioni sospette e l'impatto di nuove normative.

Domanda B6:

Quante risorse sono in staff alla funzione antiriciclaggio?

Risposta AI:

12 risorse a tempo pieno (FTE)

Source Verification

CHUNK of 2222_bancaxyz_2024_2_relval ----- Complessivamente, al 31/12/2023 la pianta organica della Funzione AML conta 12 risorse a tempo pieno (FTE) 1, di cui 1 Responsabile, 4 Analisti Senior, 5 Analisti Junior, 1 Data Scientist e 1 addetto amministrativo di supporto. Nel corso del 2023, l'organico è stato potenziato con l'inserimento di 2 nuove risorse (un analista junior e un data scientist), in

Domanda B9:

Ci sono stati cambiamenti rilevanti dell'assetto organizzativo della funzione antiriciclaggio?

Risposta AI:

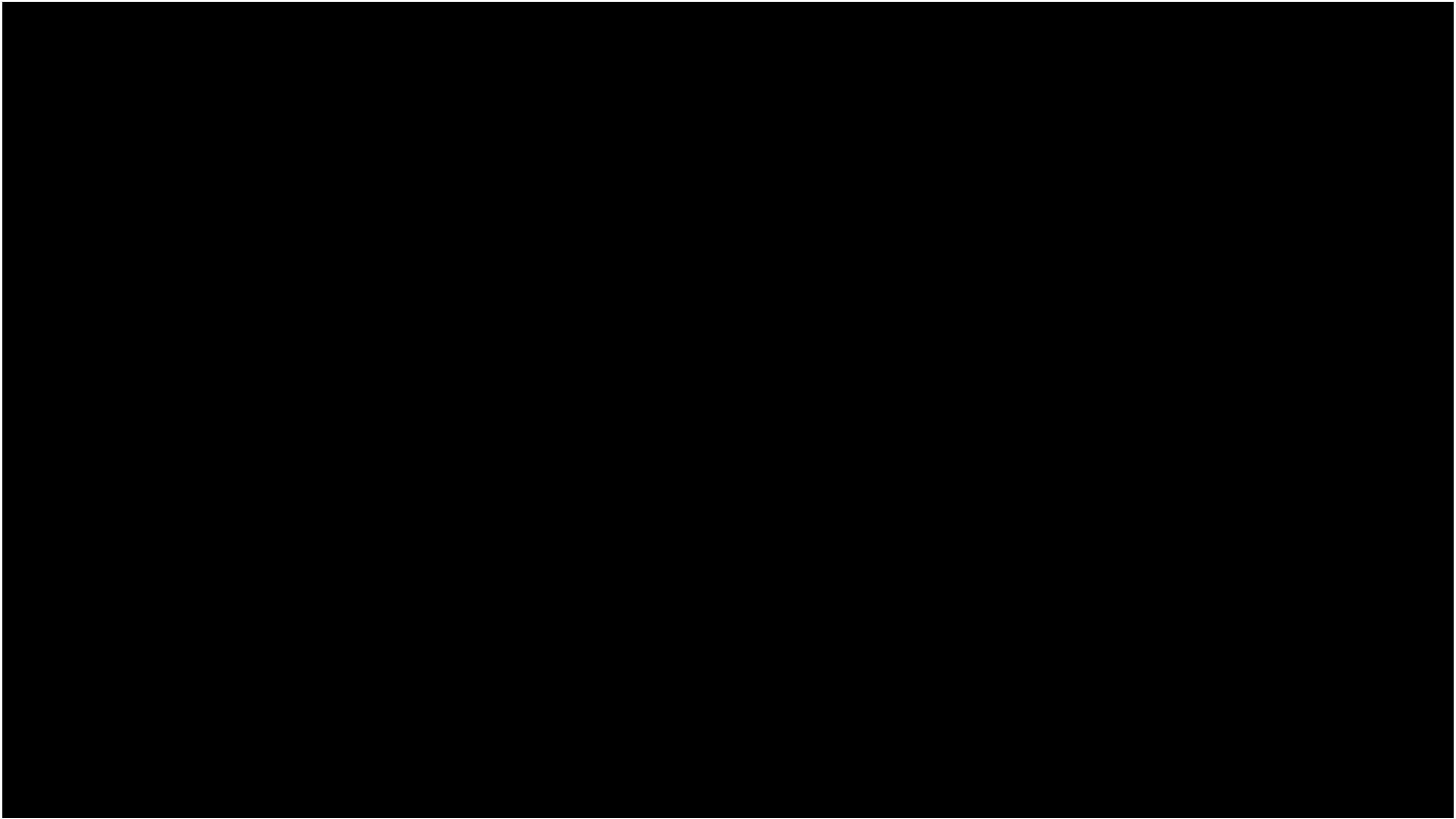
Nel corso del 2023 è stato potenziato l'organico con l'inserimento di 2 nuove risorse (un analista junior e un data scientist), in coerenza con l'aumento dei carichi di lavoro e con la volontà di rafforzare le competenze quantitative (analisi dati e modelli di rilevazione). Nel corso del 2024 è prevista l'assunzione di un analista aggiuntivo nel secondo trimestre 2024, una figura junior con competenze quantitative informatiche, che sarà dedicata al supporto nell'implementazione del nuovo sistema di monitoraggio e nell'analisi avanzata dei dati (Data Analytics AML).

Source Verification

CHUNK of 2222_bancaxyz_2024_2_relval ----- Complessivamente, al 31/12/2023 la pianta organica della Funzione AML conta 12 risorse a tempo pieno (FTE) 1, di cui 1 Responsabile, 4 Analisti Senior, 5 Analisti Junior, 1 Data Scientist e 1 addetto amministrativo di supporto. Nel corso del 2023, l'organico è stato potenziato con l'inserimento di 2 nuove risorse (un analista junior e un data scientist), in coerenza con l'aumento dei carichi di lavoro e con la volontà di rafforzare le competenze quantitative (analisi dati e modelli di rilevazione). La dotazione quali-quantitativa delle risorse è valutata adeguata rispetto alla

Every synthesized data point is traceable to the original source.

Summary Report Workflow



Results

- **Key experiment insights:**
 - Hybrid retrieval + fine-tuned embeddings improved NDCG@10 by 26% over BM25 and +15% over pretrained
 - Segmentation critical for optimal LLM performance: Best configuration with Stateless, CS 1024, Top-5*
- **Qualitative**:**
 - ~80% of generated answers useful
 - Hallucinations reduced from 3% → 0%
 - New relevant information discovered
- **Summary Report Evaluation**:**
 - RAISE vs. Copilot Chat (May 2025)
 - Criteria: readability, completeness, adherence to source data
 - RAISE outperformed baseline in **completeness & readability**, with strong source fidelity

* Tested configurations: Block vs Hierarchical segmentation; chunk sizes 1024, 2048, 3072; top-k retrieval 5, 10, 20; QA mode stateless vs stateful (20k token memory) evaluated with EM, ROUGE, BERTScore, and F1 on test set

**7 intermediaries were selected, each representing a specific type

Retrieval Performance (NDCG@10)	
Retrieval System	NDCG@10
Hybrid BM25 + bge-m3 (fine-tuned)	0.67
Hybrid BM25 + E5-small (fine-tuned)	0.64
bge-m3 (fine-tuned)	0.64
Hybrid BM25 + MiniLM (fine-tuned)	0.61
Hybrid BM25 + bge-m3 (pre-trained)	0.58
E5-small (fine-tuned)	0.60
MiniLM (fine-tuned)	0.57
bge-m3 (pre-trained)	0.55
BM25 (baseline)	0.53

Category	%	Answer Quality Breakdown**
useful for analysts	79.8	
Truly Incorrect Answers	8.3	
False Negatives	11.9	

RAISE vs. Copilot Report Evaluation**			
Criterion	RAISE	COPILOT	Δ R – C
Readability	4.17	3.66	+0.51
Completeness	3.60	2.80	+0.80
Data Adherence	3.66	3.57	+0.09
Overall Score	3.81	3.34	+0.47

Key Benefits and Constraints

- **Speed:** Reports completed in 15–20 minutes per intermediary
- **Time savings:** ~30% reduction in analysis time (≈ 400 person-days per cycle)*
- **Coverage:** Detects details not immediately obvious to analysts**
- **Traceability & Security:** Full on-premise operation, sources fully auditable
- **Constraints:** No real-time interaction, Input dependency, Annotation, Model constraints



* Based on internal analyses conducted on the reports and self-assessments of the different categories of supervised intermediaries.

** A qualitative analysis on 7 intermediaries showed an average increase of 16% (median) in information automatically extracted by RAISE compared to manually annotated data.

Conclusions

- **AI can significantly improve the efficiency** of AML/CFT document reviews.
- **RAISE is a robust and transparent system** developed entirely **on-premise**.
- Already delivered **dozens of high-quality** reports in 2025.
- **Next steps:** extension and **full operational deployment** in 2026.

THANK YOU

Reportistica Antiriciclaggio Integrata e Sintesi Estrattiva

PoC Steering Group

Role	Name	Structure
IT lead	Daniele Licari	IT Development Directorate
Business lead	Valerio Minichiello	AML Supervision and Regulation Unit
Participant	Emanuela Atripaldi	Milan - Bank Oversight Division
Participant	Daniele Bovi	IT Development Service
Participant	Andrea Da Pieve	AML Supervision and Regulation Unit
Participant	Alessio Beninati	AML Supervision and Regulation Unit
Participant	Luca Trevisan	AML Supervision and Regulation Unit
Participant	Silvia Paoluzzi	AML Supervision and Regulation Unit Milan - Asset Management And Investment Services Division
Participant	Clara Puce	