

Suspicious Activity Report (SAR) Drafting Agent

An Agentic AI Approach to TBML Detection across UK and US Jurisdictions

Project Overview

Global Capability Centres face a persistent challenge: investigations are slow, fragmented, and inconsistent as analysts manually reconcile UK and US compliance regimes. Producing a single SAR draft can take several days, straining both efficiency and quality.

Our Agentic AI model transforms this process by collapsing investigative assembly into seconds, embedding jurisdiction-specific rules directly into the workflow, and elevating analysts from data gatherers to decision makers.

We selected Trade-Based Money Laundering as the focal typology because it dominates global trade corridors, provides structured evidence trails through invoices and shipping records, and accounts for more than 60 percent of TBML cases through mis-invoicing and high-risk counterparties.

This makes it both practical to implement and strategically impactful for GCC compliance innovation.

Regulatory Landscape

UNITED KINGDOM — POCA 2002

UK POCA 2002 mandates 7 working days to respond to a consent request (DAML); if consent is refused, a 31-calendar-day moratorium period follows before the analyst may proceed.

Source: GOV.UK, Circular 008/2018 — POCA 2002 ss.335–336

UNITED STATES — BANK SECRECY ACT

This act applies to all U.S. regulated entities, regardless of the location where compliance functions are carried out. A SAR must be filed by banks for transactions aggregating USD 5,000 or more—or USD 2,000 when an identified suspect is involved—if the transaction exhibits indicators of suspicious activity.

Source: FinCEN, SAR Filing Threshold & Requirements infographic

FATF

It is not a national regulator. Its role is standard-setting. FATF's 2006 report first identified the major methods used to move illicit trade-linked funds, establishing TBML as a standing global typology.

Source: fatf-gafi.org, Trade-Based Money Laundering (2006)

Key Global Money Laundering Typologies

Trade-Based Money
Laundering (TBML)

Cyber-Enabled Fraud
& Laundering

Cash-Based Laundering

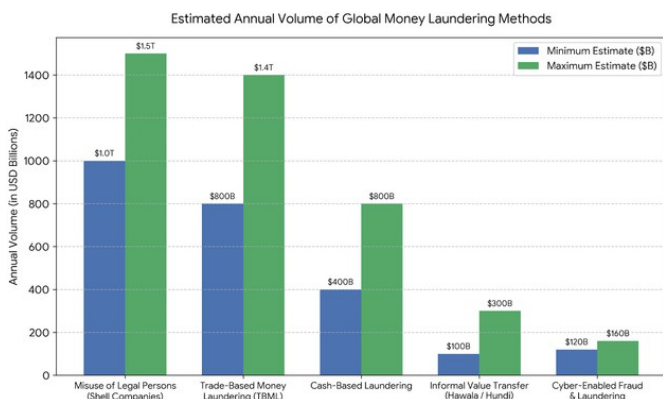
Informal Value Transfer
Systems (Hawala)

Misuse of Legal Persons
(Shell Companies)

TBML is the most significant global laundering method, with mis-invoicing alone estimated to drive over 60 percent of cases. Unlike cyber fraud or cash smuggling, it leaves structured evidence in invoices and shipping records, making it uniquely suited for AI-driven detection. Highlighting this dominance underscores why GCCs must prioritize TBML in compliance transformation. Hence it being the primary focus of this project.

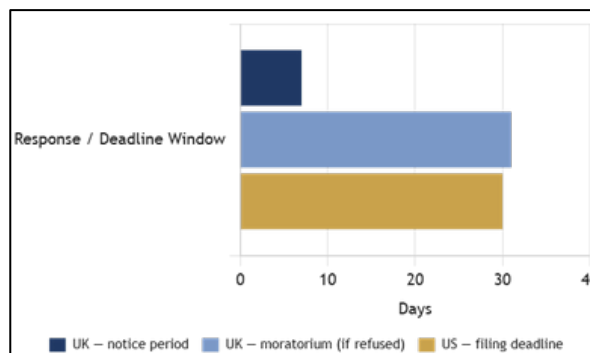
The key TBML red flag indicators include over/under-invoicing, multiple or phantom invoicing, misrepresentation of goods, high-risk counterparties, opaque payment routing, unusual trade finance instruments, shell/front companies, and transactions inconsistent with the customer profile.

FinCEN FY2023 SAR Filings by Activity Category (millions)



Source:
<https://www.undp.org/>
<https://www.trmlabs.com/reports-and-whitepapers/2026-crypto-crime-report>

Jurisdiction Comparison



Deadlines shown in calendar-equivalent days.

Source:
Sources: GOV.UK Circular 008/2018;

Jurisdictional Framework Comparison

Dimension

United Kingdom

Legal Framework

POCA 2002, Part VII; Terrorism Act 2000

Filing Authority

National Crime Agency (NCA UKFIU)

Filing Threshold

No monetary threshold — any suspicion

Deadline

7 working days for NCA response; 31-day moratorium if refused

SAR Format

NCA Online Disclosure (free-text)

Penalty

MLRO: up to 5 years imprisonment

United States

Bank Secrecy Act (1970); 31 CFR Part 1020

FinCEN

USD 5,000 (USD 2,000 for MSBs)

30 days from detection; 60 if no suspect

FinCEN Form 111, Section C narrative

Civil up to \$25,000/day; criminal up to 10 years

Why Is This Agentic AI?

Unlike a fixed single-path workflow, information branches only at runtime — jurisdiction is unknown until the profile is read; pass/fail of quality checks is unknown until the LLM drafts and checks run.

The LLM never issues final approval — a human does. A rejected draft routes back to the drafting node rather than terminating.

The system self-corrects by feeding specific failure reasons back into the LLM prompt rather than blindly retrying.

This is **bounded agentic AI** — the agent chooses among pre-approved, compliant paths rather than inventing new ones, appropriate for a regulated AML context.

Tool Integration

The agent is built on **LangGraph**, which orchestrates a typed **SARState** object carrying case data, risk score, jurisdiction, draft, and revision count across every node. An external **LLM API** is called only for narrative generation and re-drafting — all detection, routing, and validation logic is deterministic code, not model output.

Deterministic routing + generative drafting: rules decide *what* to file, the LLM decides *how to phrase it*.

A single typology library feeds both the risk-scoring node and the narrative generator, keeping detection and drafting logically consistent. The review UI surfaces exactly the fields the diff-check couldn't verify, so analyst attention goes where it's needed.

Sources:
<https://phase2online.com/insights/understanding-langgraph-for-llm-powered-workflows>
<https://timjwilliams.medium.com/llm-verification-loops-best-practices-and-patterns-07541c854fd8>

Operational Impact

SCALE DRIVER

Instead of analysts spending 2–5 days piecing together fragmented data and drafting SARs from scratch, the agent collapses the entire investigative assembly into seconds. By parallelizing KYC aggregation, typology matching, jurisdiction routing, and narrative generation, it shifts human effort from manual writing to high-value judgment and MLRO sign-off. The result is a step-change in compliance operations: speed at scale, uniformity across teams, and uplifted quality through confidence-scored narratives that spotlight uncertainty for targeted scrutiny.

This version positions the system not just as an efficiency tool, but as a transformational re-architecture of AML workflows.

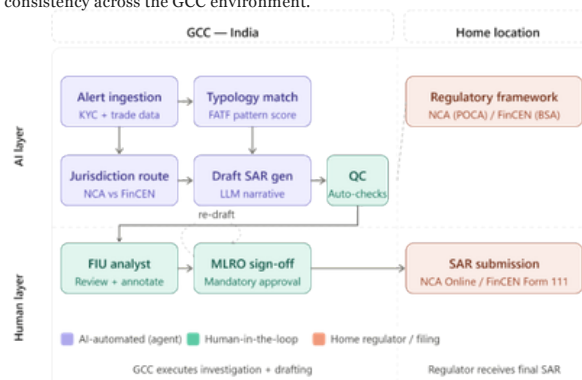
The role of human analyst is the finalisation and approval of the draft generated by the agentic AI.

BENEFITS

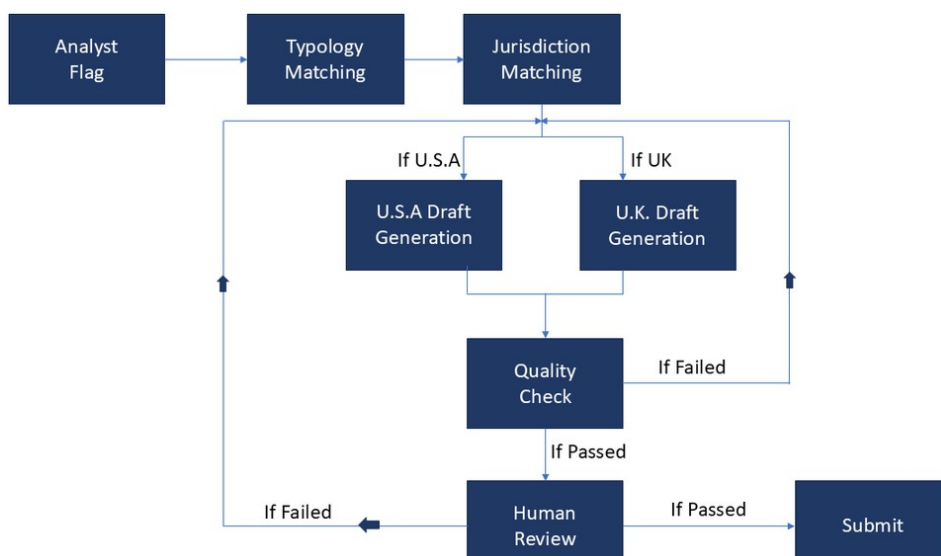
The system delivers speed compression by reducing processes from days to seconds, ensures consistency across analyst teams regardless of tenure, and drives quality uplift through confidence-scored narratives that highlight low-certainty fields for targeted human scrutiny.

GCC RELEVANCE

GCCs operate under a dual-accountability model, based in one jurisdiction while delivering compliance outcomes worldwide. Analysts must apply both UK and US regimes, each with different thresholds, deadlines, and formats. Our agent eliminates the need for every analyst to master both regimes, ensuring speed, accuracy, and consistency across the GCC environment.



Agentic AI Workflow



This workflow diagram illustrates how flagged cases move through typology and jurisdiction matching, ensuring draft generation aligns with U.S. or U.K. regulatory logic. Quality checks and human review act as control gates, looping back on failures to maintain accuracy and compliance.

The structured flow culminates in submission, embedding consistency and accountability across the SAR process.

SAR FILING AGENT

AML COMPLIANCE SYSTEM - TML DETECTION

Developed by Students of IM

NEW CASE

Customer ID

CUST-002 — Pacific Rim ...

Analyst ID

4 — Gaurav Singh

Running...

CUSTOMER PROFILE

Name

Pacific Rim Traders Inc

Country

United Kingdom

Business

Trading Company

KYC Tier

Tier 2

PIPELINE STATUS

Information Retrieval

Typology Matching

Jurisdiction Routing

Draft Generation

Quality Checks

Human Review

Live Output

Analyst Review

Result

MANDATORY ANALYST REVIEW — MLRO SIGN-OFF REQUIRED

CASE ID

SAR-202607-FAF408

CUSTOMER

Pacific Rim Traders Inc

REGULATOR

FinCEN

RISK SCORE

0.90

JURISDICTION

United Kingdom

FLAGGED TXN

14

FLAGGED TRANSACTIONS

transaction_id	date	invoice_date	commodity	quantity	unit_price	invoiced_amount	fair_market_value	fair_unit
TXN-0001	13-Jan-2024	11-Jan-2024	Textile & Apparel	54	27.11	1463.7	5625.27	104.17
TXN-0009	20-Jan-2024	15-Jan-2024	Electronic Components	67	1545.62	103556.7	36571.9	545.85
TXN-0017	27-Mar-2023	20-Mar-2023	Electronic Components	256	474.85	121561.13	35390.09	138.24
TXN-0025	26-Oct-2023	17-Oct-2023	Electronic Components	230	140.55	32327.59	12698.7	55.21
TXN-0033	28-Mar-2023	26-Mar-2023	Construction Materials	273	190.45	51994.03	21020.72	77
TXN-0041	04-Apr-2023	03-Apr-2023	Electronic Components	204	11.92	2432.69	11525.79	56.5
TXN-0049	31-Dec-2023	23-Dec-2023	Auto Parts	415	97.94	40645.37	14260.13	34.36
TXN-0057	26-Apr-2023	23-Apr-2023	Pharmaceutical Inputs	433	395.41	171211.42	74673.77	172.46
TXN-0065	03-Oct-2023	25-Sep-2023	Agricultural Commodities	180	51.65	9297.12	17237.09	95.76
TXN-0073	12-Sep-2023	07-Sep-2023	Electronic Components	320	328.84	105228.38	31769.02	99.28

Risks	Mitigations
Hallucinated Narratives	Quality Check against the source transaction dataset
Jurisdiction Routing Errors	Jurisdiction Identification node, Jurisdiction Specific Drafting Nodes
False Negatives on Novel Patterns	Human Review Node
Data Privacy	Self Hosted Local LLM

Roadmap

Our roadmap extends beyond the proof of concept.

We aim to migrate from server-based APIs to self-hosted models for greater data privacy, integrate live core-banking and trade-finance datasets, and expand coverage to all eight TBML red flags.

Future iterations will embed analyst feedback loops into risk scoring, enabling continuous learning and adaptation.

This positions the agent not just as a workflow accelerator, but as a scalable compliance transformation engine.