

AI Governance in Irish Financial Crime



Author: David Kearney, experienced PCF 52 and Member of Compliance Institute’s Financial Crime Compliance Working Group.

Governance of Advanced Financial Crime Compliance Related Artificial Intelligence and Agentic Systems in Irish Regulated Firms

The Irish financial services landscape is currently navigating a period of profound structural transformation, driven by the convergence of heightened regulatory expectations and the rapid integration of advanced artificial intelligence. As firms transition from traditional machine learning models to autonomous agentic frameworks and large-scale learning models within their financial crime compliance functions, the requirements for robust governance have moved from the periphery of technical oversight to the center of compliance risk management.

The Central Bank of Ireland (CBI) has explicitly identified technology-driven change as a primary supervisory priority for 2026. The adoption of advanced AI in Financial Crime compliance is no longer merely an innovation initiative but a fundamental component of a firm’s operational resilience and individual accountability obligations.

The Evolving Regulatory Landscape and Supervisory Expectations

The regulatory environment for Irish-regulated firms increasingly focuses on how technology influences the safety and soundness of the financial system. The CBI’s Regulatory and Supervisory Outlook for 2026 establishes a clear mandate for firms to demonstrate that their adoption of AI does not outpace their capacity to govern it. This supervisory focus is harmonized with broader European frameworks, including the [Digital Operational Resilience Act \(DORA\)](#) and the EU AI Act, creating a complex web of requirements that firms must navigate to maintain compliance and public trust. The CBI has identified five industry-wide priorities for the year ahead, with a particular emphasis on responding to technology-driven change and protecting consumer interests in a digital financial system. For firms utilizing AI in financial crime compliance, this means moving beyond isolated compliance exercises toward an integrated, outcomes-based governance model. The CBI expects boards and senior management to be able to evidence that they understand the deployment of AI within their organizations and can provide effective challenges to automated decisions.

CBI Supervisory Priority Area	Specific Impact on Financial Crime Compliance	Governance Expectation for AI Systems
Operational Resilience	Ensuring AI-driven transaction monitoring remains stable during disruptions.	Integration of “kill switches” and manual failover protocols for autonomous agents.
Technology Risk	Managing the cybersecurity and dependency risks associated with AI pipelines.	Strict adherence to ICT risk management frameworks under DORA.
Consumer Protection	Preventing algorithmic bias in automated AML risk profiling and client onboarding.	Rigorous fairness auditing and transparent explainability of decision-making.
Financial Crime	Using AI to counter sophisticated social engineering, deepfakes, and payment fraud.	Deployment of real-time monitoring and advanced pattern recognition capabilities.
Individual Accountability	Mapping algorithmic oversight to senior executive responsibilities under SEAR.	Clear documentation of “reasonable steps” taken to oversee AI performance.

The CBI's message is consistent across sectors: capital strength is no longer the sole metric of institutional health; regulators are now examining how firms manage technology risk and operational disruption in practice. For compliance professionals, The governance framework surrounding the development of advanced AI tools is becoming a critical success factor

Establishing a Robust Governance and Control Framework

The transition from traditional analytical AI to “Agentic AI” represents the next evolutionary step in financial crime compliance. While traditional machine learning models have been used for years to score alerts and reduce false positives, agentic systems possess the ability to carry out tasks and make decisions autonomously with minimal human intervention.

Agentic AI differs from Generative AI (GenAI) in its focus on action rather than just content creation. In a financial crime context, an AI agent can analyze transactions against risk thresholds in real-time, reason through anomalies, retrieve relevant customer data from multiple internal and external databases and autonomously decide whether to clear an alert or escalate it for human review.

The primary governance challenge associated with autonomous agents is “behavioural safety”. Unlike traditional software that follows static linear rules, agentic systems are probabilistic. They evaluate situations and select actions based on non-linear reasoning, which can sometimes lead to scenarios where the logic behind a specific decision becomes opaque or unaccountable.

For Irish regulated firms, the development of an appropriate control framework for advanced AI is an operational imperative. This framework must be integrated into existing enterprise risk management (ERM) structures and designed to address the unique challenges of model drift, bias, and autonomous action.

Traditional model validation frameworks require significant enhancement to cater to the complexities of AI and machine learning. A modern Model Risk Management (MRM) is a structured, enterprise-wide framework used to identify, measure, monitor, and mitigate risks—such as financial loss, poor decision-making, or reputational damage—arising from the design, development, or use of inaccurate or misused quantitative models. Just drawn from AI, sure you have a better/more tailored one.)

Firms must ensure that every model is “fit for purpose”, balancing the interrelated dimensions of model fairness and model accuracy.

MRM Dimension	AI-Specific Challenge	Required Validation Activity
Data Quality	Training data may contain historical biases or be unrepresentative.	Documented assessments of bias, representativeness, and formal sign-off of datasets.
Explainability	Advanced models (deep learning) can be inherently opaque.	Use of interpretability tools (e.g., SHAP, LIME) to map decision drivers.
Model Drift	Accuracy degrades as criminal tactics and economic conditions change.	Real-time monitoring of feature distribution and ongoing back-testing.
Bias and Fairness	Algorithms may unintentionally discriminate against protected groups.	Periodic auditing of outcomes against gender, religion, and race metrics.
Operational Stability	Autonomous agents may malfunction or reach “hallucinated” conclusions.	Rigorous stress testing and “self-heal” verification for agentic workflows.

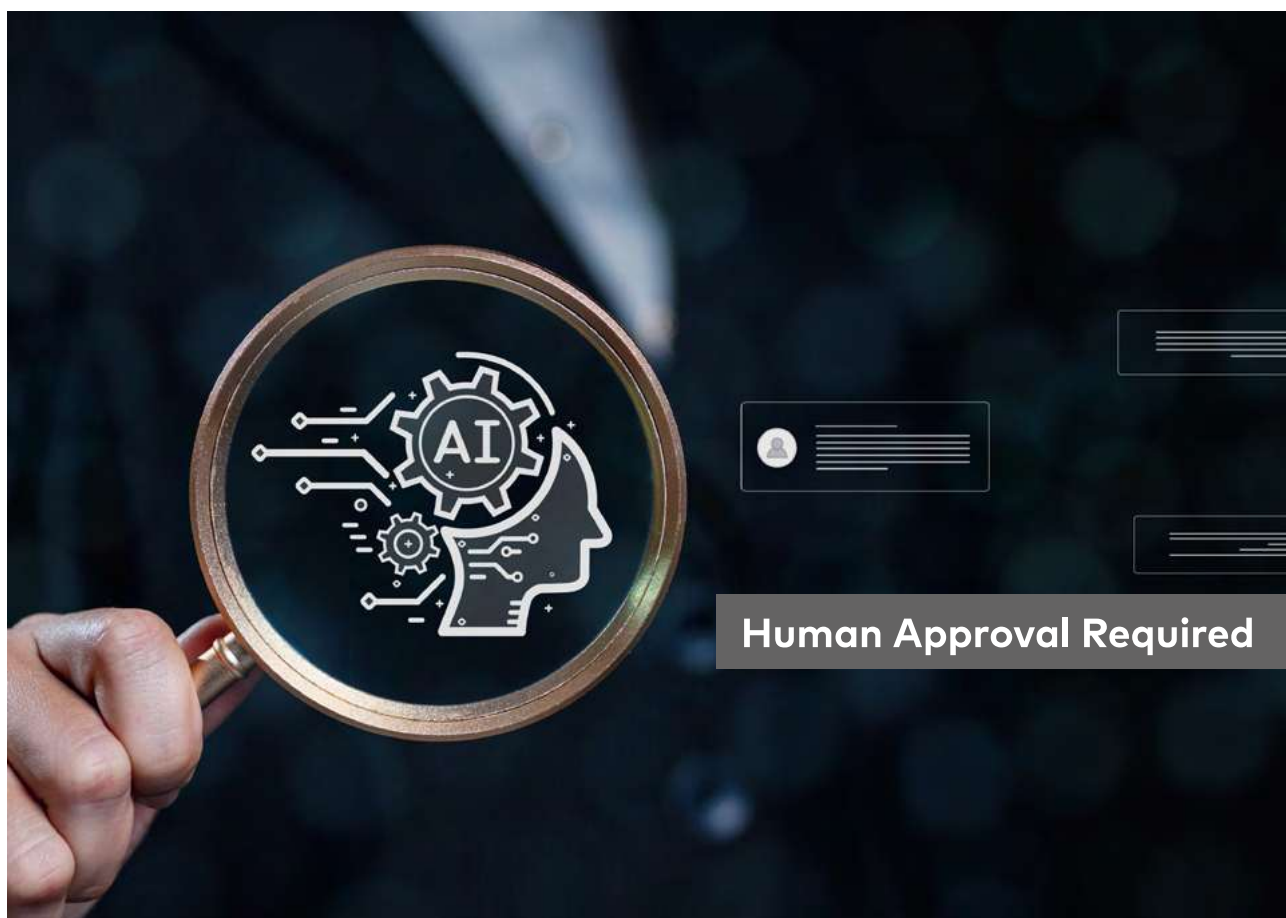
expect firms to be able to “show their program” and explain how an algorithm filtered a million financial crime alerts without missing a known criminal or a sanctioned individual.

Effective AI governance should follow a proportionate, risk-based approach. Not every AI deployment requires the same level of control; while a chatbot used for general employee research might operate with limited oversight, AI tools that influence production decisions—such as setting

AML risk thresholds or clearing transaction alerts—require mandatory validation and monitoring.

Firms may implement a “Tiered Autonomy” model:

- **Human-in-the-loop (HITL):** For high-risk financial transactions or finalizing SARs, explicit human approval must be required before an action is executed. The AI agent acts purely as an advisor.
- **Human-on-the-loop (HOTL):** For mid-risk tasks



like adverse media summarization, AI agents may act autonomously within set parameters, with humans monitoring in real-time to intervene or use a “kill switch” if needed.

- **Human-post-action audit:** For low-risk content generation like horizon scanning a retroactive audit of outputs may be sufficient to ensure ongoing quality and compliance.

This tiered approach ensures that human accountability is preserved and an ethical approach applied where it matters most, preventing the “accountability gap” that can occur when autonomous systems fail.

Implementation Strategy: Moving from Innovation to Integration

To successfully operationalize advanced AI agents in the current regulatory climate, Irish firms should adopt a structured, phased implementation strategy that prioritizes governance and cultural alignment.

Success in AI-led compliance requires more than just technical tools; it requires a workforce that is “AI literate”.

Firms must ensure that AI literacy training is tailored to staff roles, risks and technical knowledge

focusing on understanding AI functioning, risks and ethical usage. A digital-first culture, as advocated by the EBA, involves moving beyond ad hoc controls to a holistic governance framework. This culture should encourage continuous learning and “effective challenge,” where compliance staff are empowered to question and override AI-driven recommendations.

To meet the auditability requirements of the EU AI Act and the transparency expectations of the CBI, firms should implement an automated “flight recorder” for their AI environment. This system should track every interaction with sensitive datasets and every decision made by an AI agent, comparing activity to compliance policies in real-time. This always-on-oversight transforms compliance from a reactive, quarterly review process to a proactive, real-time risk management discipline. It ensures that when a regulator asks, “Why did the AI agent clear this alert?”, the firm can pull an explainable, documented record of the reasoning path used.

Future Outlook

As we look toward to 2027, the role of AI in financial crime compliance will only deepen. The European Anti-Money Laundering Authority (AMLA) is expected to be fully operational by 2028, bringing



a common supervisory methodology that will likely incorporate its own supervisory technology (SupTech) tools to enhance oversight of firms. (Super Tech is a term used to describe a suite of innovative tools and technologies which are being developed by component authorities to address AML/CFT supervision tasks and process relating to risk assessment and the management of data.)

The Consumer Protection Code 2025 (CPC) became effective in March 2026 and places greater emphasis on “digital service delivery” and the treatment of vulnerable customers which are key areas of focus for the CBI. Firms must ensure that their AI agents are designed to recognize and appropriately handle the needs of vulnerable customers, avoiding “algorithmic exclusion” where legitimate users are barred from services due to rigid or biased automated profiling.

The 2026 regulatory landscape in Ireland will not penalize firms for using AI, but it will certainly penalize those that use it without a

clear, documented, and accountable governance framework. The future of compliance is digital, but it remains fundamentally human in its accountability and ethical application.