

AI & ML

CASESTUDY

Trust by design:
scaling AI without the regulatory hangover

Background

A specialised lender (“the Client”) in the Australian car-loan market and recently expanded into digital savings products across the EU and US via passportable e-money licences and a US ILC charter application. Rapid adoption of AI-driven credit scoring, chat-bots, and marketing models triggered board concern after regulators signalled tougher scrutiny under the EU AI Act (2026 full application), US executive-order mandates on AI safety, and APRA’s draft CPS 230 tech-risk standards. The bank’s legacy model-risk policy covered only traditional logistic-regression scorecards—leaving LLMs, reinforcement-learning optimisers, and embedded third-party APIs largely unmanaged.

Introduction

Enigma Risk Advisory was engaged to architect an enterprise-wide AI Risk Management & Responsible-AI Framework. The remit:

- Map all AI uses: commercial SaaS (e.g., GPT-4o-Enterprise), on-prem OSC LLMs (Llama 3 derivatives), and in-house ML pipelines—across the model lifecycle.
- Align with EU AI Act, US National Institute of Standards and Technology (NIST) AI Risk Management (RMF), Australia’s Safe & Responsible AI Principles, and Digital Operational Resilience Act (DORA) ICT-risk obligations.
- Deliver auditable, developer-friendly controls without throttling product velocity.

Challenges Faced

Theme	Issue	Risk Consequence
Model Development	27 AI models built in three languages with no unified metadata; shadow models in business units	Opaque lineage → replication failures, audit fines
LLMs (Commercial vs. Open Source)	Commercial GPT-4o for chat-bots, open-source Llama-3-70B-finetune for credit underwriting	Inconsistent governance, data-export risk to US cloud, licence ambiguity
Regulatory Patchwork	EU AI Act “high-risk credit-scoring” rules, US CFPB fair-lending AI guidance, APRA CPS 230 + draft AI position paper	Conflicting timelines → compliance gaps
Privacy & IP	PII ingestion into prompt payloads; vendor terms allow derivative-model training	GDPR and CCPA breach exposure; IP leakage
Operational Resilience (DORA)	AI inference latency sits on single-region GPU cluster	Concentration & ICT-outage risk; resilience testing gaps



Results and Impact

- 100% AI inventory visibility—all models registered with lineage, licences, and risk tier within six weeks.
- Model validation cycle-time ↓ 40% via automated test harness (bias, drift, adversarial robustness).
- No critical findings in joint ECB-APRA supervisory review; DORA ICT-risk Self-Assessment score improved from 58% → 92%.
- Fair-lending rejection-gap narrowed 18% for protected classes after bias-mitigation retrain.
- \$3.7m annual cost avoided by rationalising overlapping SaaS LLM usage; negotiated vendor T&Cs ensure no training on bank data.

Solutions Implemented

a. AI Governance Blueprint

- Three-Lines-Plus: Product Teams → Model-Risk Engineering (independent code review) → Compliance & Legal → Board AI & Ethics Committee.
- Risk-tier matrix aligned with EU AI Act risk classes; 11 mandatory controls for “high-risk” models (credit-scoring, AML).

b. Unified ModelOps Platform

- Open-source MLflow + custom metadata layer; CI/CD enforces 4-Eye peer review, explainability artefacts, FAIR data tags.
- Policy-as-Code: YAML guardrails insert PII redaction, licence checks, and prompt-logging before production deploy.

c. LLM Strategy & Segmentation

Use Case	Engine	Governance Action
Customer chat-bot	GPT-4o-Enterprise (Azure EU)	Contractual data-isolation clause + red-team jailbreak tests
Credit-underwriting assistant	Llama 3-70B-finetune (on-prem)	Secure enclave, differential-privacy fine-tune, SOC 2 Type II review
Marketing copy gen	Mixtral 8×22B (API)	Low-risk tier; watermarking + human-in-the-loop

d. Regulatory & Legal Alignment

- Cross-jurisdiction compliance map: EU (AI Act, GDPR, DORA), US (CFPB fair-lending, EO 14110), AU (Safe AI Principles, CPS 230).
- Automated template generates AI System Record for each model—cites harmonised legal bases, DPIA results, and DORA resilience tests.

e. Resilience Engineering (DORA)

- Multi-cloud GPU failover; chaos-engine rules inject latency faults weekly.
- Scenario bank covers data-poisoning, vendor LLM outage, and sovereign-cloud exit.

f. Cultural & Training Uplift

- 200+ staff certified through “Responsible-AI Bootcamp”; #AI-check Slack bot flags unregistered code snippets.

Conclusion

By embedding governance directly in code pipelines, contract clauses, and ICT-resilience playbooks, the Client now deploys AI at scale. The framework balances innovation speed with ethical-legal rigour, positioning the bank as a regional leader in trustworthy AI while protecting customers—and balance-sheet—from emerging algorithmic risks.