MINDFORGE

AI Risk Management: Implementation Examples

January 2026

MINDFORGE

The MindForge Consortium

MAS Monetary Authority of Singapore

AIA HEALTHIER, LONGER, BETTER LIVES · BlackRock · citi · DBS · eastspring investments A Prudential plc company

Great Eastern An OCBC Company · GIC · GXS · HSBC · HSBC Life

income made yours · Julius Bär · Manulife · Maybank · MSIG

MUFG · Munich RE · OCBC · PRUDENTIAL · SMBC

standard chartered · STATE STREET · UBS · UOB

aws · Google Cloud · Microsoft · NVIDIA

Supported by: accenture

With participation from:

abs The Association of Banks in Singapore · General Insurance Association · imas Investment Management Association of Singapore · Life Insurance Association Singapore Life is worth protecting. Invest in it. · SFA Singapore FinTech Association

Contents

Introduction

Financial institutions (FIs) are actively working to manage the risks of artificial intelligence (AI) in their organisations while also taking advantage of its benefits for their customers, employees, and communities. The experience gained by these FIs has been instrumental in the development of the AI Risk Management Handbook and serves as a useful point of reference for others in the sector who look to turn the Handbook's considerations into real-world practice.

This document features experiences and lessons learned from four FIs at different levels of maturity as they have implemented AI risk management in their organisations. These FIs are, in alphabetical order:

- DBS
- Julius Baer
- Prudential
- Investment Firm

The consortium sincerely thanks these four FIs for their effort and enthusiasm in contributing implementation examples as part of this Handbook.

The Handbook consists of three documents, of which this is the third. These three documents are:

AI Risk Management Executive Handbook. This document provides Considerations and Implementation Practices for governing AI across each Section in the Handbook's scope. It is intended as a resource for executives in the financial services industry.

AI Risk Management Operationalisation Handbook. This document provides detailed guidance on the operationalisation of each of the Implementation Practices recommended under each of the Handbook's Consideration. It includes illustrations of good practices from primary members, appendices, and other supporting materials.

AI Risk Management Handbook Implementation Examples (This Document). This document provides detailed case studies on individual financial institutions' experiences implementing AI governance and risk management.

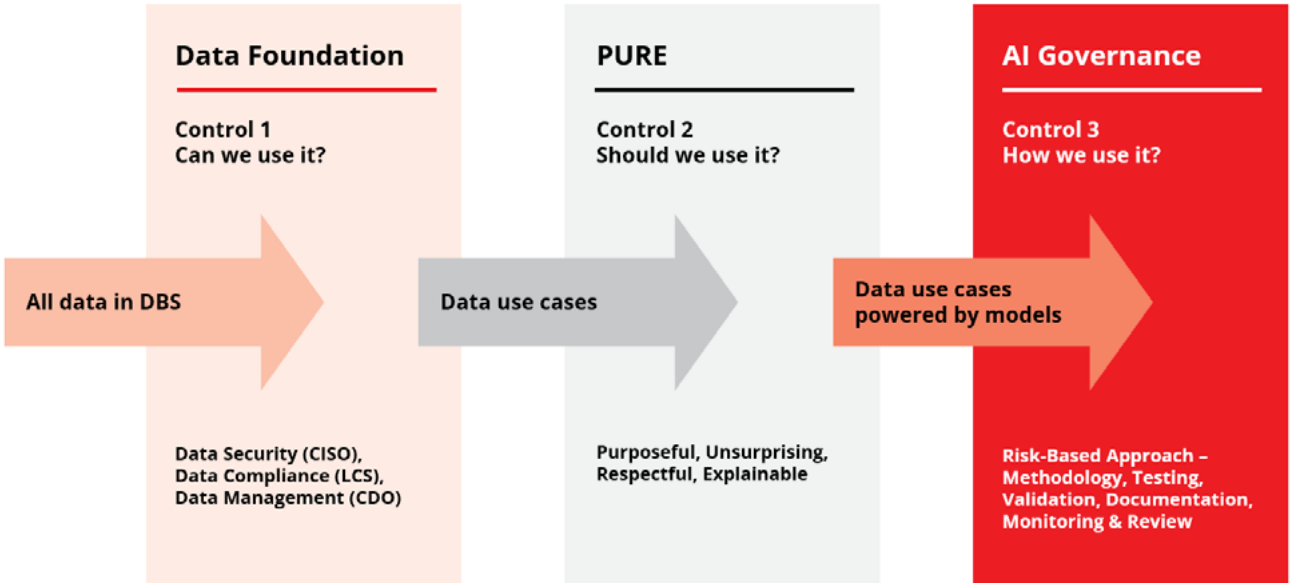
These three documents are meant to be used in conjunction, and together make up the Handbook.

DBS Implementation Example

DBS Implementation Example

DBS' Responsible Data Use (RDU) Framework

At DBS, we recognise that the responsible use of data is increasingly essential as we expand the scale and pervasiveness of AI across the bank. As we continue to explore the potential of AI, we remain steadfastly cautious of its associated risks. Our **Responsible Data Use (RDU) framework** helps to ensure our data usage and AI adoption are lawful, ethical, and fair by addressing three core questions:



Data Foundation – Can we use it?

We established a data policy framework to foster a robust approach to foundational data management aspects such as data security, privacy, access and quality. This ensures that our data is securely managed, fit for the intended use, and strictly adheres to internal standards and relevant laws and regulations from the outset.

DBS PURE framework – Should we use it?

To address the ethical dimensions of our data use in alignment with DBS' core values, the DBS PURE framework guides the use of data in the specific business context, emphasizing Purposeful, Unsurprising, Respectful, and Explainable (PURE) data use. The PURE framework serves as the bank's ethical compass and has been in use since 2019. It is regularly enhanced to uphold the PURE principles as we tap on the potential of data and AI amid evolving regulatory regulations, customer expectations, and societal norms.

AI Governance – How we use it?

We are dedicated to robust governance practices in the development and deployment of AI systems, ensuring fairness, transparency, interpretability, and accountability in our use of AI. Our risk-based approach to AI governance ensures appropriate and proportionate governance across the end-to-end lifecycle of AI use cases. Recognising the complexities introduced by Gen AI, we also implemented additional guardrails and controls to safely explore, test, and adopt these emerging and rapidly evolving technologies.

Operationalising AI Governance

To scale the use of AI across the bank, our **Data & AI Industrialisation Programme** drives a strong data-driven strategy through these pillars:

1. **Technology:** A centralised data and AI platform **(ADA)** with a modern hybrid architecture and advanced analytics capabilities.
2. **Process:** Our AI protocol **(ALAN)** provides standardised processes and best practices, serving as a central knowledge repository for all AI use cases in the bank.
3. **People:** The Data Chapter **(DC)** fosters talent effectiveness, bringing together 700 Data Professionals (Analysts, Translators, Scientists) to deliver business outcomes at speed and at scale.

Our adoption of AI is further underpinned by an AI governance framework, structured around 5 key features, that drives robust oversight and accountability to sustain confidence and trust in AI outcomes:

- **AI Risk Materiality Assessment:** The risk materiality of each AI use case is determined against materiality rubrics defined and endorsed by the RDU Committee.
- **Calibrated Governance Requirements:** Baseline or enhanced requirements are applied based on the use case's risk materiality.
- **Central AI Repository:** All AI use cases and underlying models are registered in ALAN, which helps orchestrate the application of AI governance requirements across the AI lifecycle.
- **Operating Model:** Clearly defined roles and responsibilities are assigned within each unit and location to support AI governance activities.
- **Senior Management Accountability:** Oversight by Senior Management and the Board of Directors drives clear awareness and accountability for AI governance.

End-to-End AI Governance on CodeBuddy

In recent years, the use of AI models to solve business problems at DBS has increased substantially. To meet this increased demand while optimising scarce data science and analyst resources, the bank leveraged Gen AI to boost the productivity of its DC employees.

CodeBuddy, an in-house Gen AI-powered programming assistant, streamlines coding tasks for analytics and AI/ML to accelerate the delivery of data-driven insights and models. It integrates large language model (LLM) capabilities with ADA and DBS-specific coding knowledge to provide up-to-date and contextually-relevant code assistance, offering features like code completion, generation, explanation, debugging, and question answering. This reduces the time spent on data exploration, analysis, feature creation, and model training, thereby ensuring the quality of the models, managing risks, and accelerating the onboarding and productivity of new analytics talent.

Since its inception, over 80% of our target DC employees actively utilise CodeBuddy's capabilities, with positive results supported by the responsible AI practices applied across the AI lifecycle.

Use case context & design:

The development of CodeBuddy began with the clear identification of use case and model owners from the outset to ensure end-to-end accountability across the AI lifecycle. A preliminary risk materiality assessment, approved by senior management prior to deployment, evaluated the solution's potential adverse impact and autonomy to identify appropriate governance in line with organisational standards on AI governance. To ensure transparency and effective governance across the lifecycle, use case and model details were documented in ALAN from the early stages. CodeBuddy also employs a human-in-the-loop approach, requiring active user review and approval of generated outputs to mitigate risks from AI inaccuracies or unintended behaviour.

Data acquisition and processing:

Robust data management controls ensured ethical and lawful data handling that is aligned with DBS' core values. This includes a detailed PURE assessment and uplifting security controls in ADA, such as the implementation of stateless processing, for LLMs to safely interact with DBS data and prevent unauthorised third-party access. Building CodeBuddy within the ADA environment allowed it to leverage the security and data access control mechanisms inherent in the platform. Although CodeBuddy functions as a coding assistant, users (DBS employees) bear full accountability for verifying the correctness and suitability of the generated code. The use of retrieval augmented generation, which integrates the pre-trained LLM with our internal knowledge base, grounds code suggestions in verified sources so that the risk of errors or potential biases in the model's pretrained data are mitigated.

Onboarding, build, and review:

During LLM onboarding, third-party risk profiling was conducted to manage risks associated with third-party AI technology. This ensures that only approved LLMs are permitted for use within DBS, and that they are accessible via a secured proxy service for governed interaction with these external products. Post-onboarding, rigorous performance testing was conducted to assess the LLM's ability to perform the core functions required of CodeBuddy and mitigate potential risks such as coding inaccuracies and hallucinations. A key transparency measure integral to its human-in-the-loop design is the prominent disclaimers built within the user interface, reminding users to verify the outputs generated.

Deployment:

Prior to deployment, a comprehensive monitoring plan was established for CodeBuddy, outlining key performance metrics, reporting frequency, acceptable performance thresholds, and communication/ feedback channels to ensure continued fitness for use. A contingency measure which allows users to disable the coding assistant feature was also implemented to prevent disruption to development activities. The use case then underwent a thorough review and approval by the cross-functional RDU Committee for a holistic assessment of the potential Gen AI risks. To further validate its performance and mitigate unforeseen risks, a progressive, phased rollout was implemented. Trainings were also conducted to equip users with the necessary competencies for proficient and responsible use of the coding assistant.

Usage, monitoring, and change management:

Post-deployment, CodeBuddy underwent continuous monitoring to track performance against defined metrics, assess the adoption rates and collect user feedback. This facilitates proactive interventions and identifies areas of improvement. In line with our AI change management process, changes (e.g. LLM upgrades) are thoroughly tested before production deployment to mitigate potential risks arising from such changes. Such changes are also documented, reviewed, and approved to ensure traceability and accountability. Additionally, major changes are subject to peer review processes, as well as any necessary communication to downstream users.

Overcoming AI Challenges

DBS had to overcome several key challenges in its journey to industrialise responsible AI:

- **Fostering a culture of responsible AI through employee education:** Scaling responsible AI use requires employees to have a strong foundational understanding of data and AI. Since 2019, DBS has developed extensive novice and practitioner training modules on our DigiFY platform to build data management awareness and capabilities, with over 126,000 modules completed to date. To foster continuous learning in this evolving landscape, the bank also introduced a "Fairness" practitioner module for ethical AI development, integrated a 'PURE' learning component in our new joiner curriculum to underscore the importance of responsible data use to all employees, and hosted a "Gen AI Unplugged" webinar series featuring external experts on responsible AI practices and emerging challenges.
- **Integration into legacy systems and workflows:** Incorporating AI into existing banking operations can be complex and resource-intensive. We address this by leveraging ADA to support modular integrations, reusable model components, templates, and automation. This has significantly reduced AI project timelines from 15 months to under 3 months.
- **Addressing incremental risks with adoption of new AI technology:** It is essential to continuously strengthen our AI governance practices to address incremental risks introduced by rapidly evolving AI. For example, our initial scope of Gen AI adoption was intentionally designed for internal use with high levels of human oversight and incremental adoption. We also established a cross-functional RAI taskforce comprising senior and experienced subject matter experts from core functions to thoroughly evaluate use case pilots and guide risk mitigation. This taskforce, coupled with elevated clearance through the RDU Committee, ensures sufficient senior management oversight on Gen AI use cases.

Shaping Future AI Governance Efforts

Our journey in AI governance yielded important lessons that shape our future efforts:

- **Multidisciplinary collaboration:** DBS' approach to responsible AI is built upon extensive collaboration with internal stakeholders from the business, analytics, risk, compliance, technology, and human resource functions. Continuous engagement with regulators and industry bodies (such as through Project MindForge) further drives collective progress in AI governance within the financial services industry.
- **Adaptive governance:** The rapid evolution of AI necessitates an adaptive governance approach, with the need to review and enhance existing governance practices to ensure that they remain relevant and effective over time.
- **Proportionate and risk-based approach:** Proportionate governance efforts based on risk materiality are crucial for efficient risk management without stifling innovation.
- **Continuous learning:** A continuous learning culture enables the organisation to keep pace with evolving AI technology, regulations, and societal norms.

DBS continues to work closely with regulators and industry bodies to progress AI governance knowledge in the community as we strive towards becoming a leader in responsible AI use.

Julius Bär

Julius Baer Implementation Example

Julius Bär

Julius Baer Implementation Example

Enhancing Employee Productivity with a Governed Generative AI Solution: The General Chatbot Use Case

Introduction: Julius Baer's AI Governance Landscape

At Julius Baer, AI is integrated into operations under a robust governance and independent validation framework aligned with regulatory expectations and industry best practices. This approach ensures that innovation proceeds safely, particularly when handling sensitive data.

The AI governance process comprises three main stages prior to deployment – business prioritisation, review against AI regulations, and risk assessment (including validation) – followed by ongoing monitoring throughout the lifecycle of each use case.

Business prioritisation begins with an assessment of business value. Gen AI initiatives are evaluated centrally to encourage reuse, efficient resourcing, and adequate support for citizen development, while traditional AI projects are prioritised within individual business areas.

During the review against AI regulations, proposals are screened against applicable regulations, including the EU AI Act. This step identifies and excludes any use cases that may fall into prohibited categories. Only those passing this checkpoint proceed to development.

The next phase involves an assessment conducted by designated risk owners. To accommodate AI-specific risks, existing risk frameworks have been augmented with AI-related dimensions. This integration ensures comprehensive evaluation by specialists in areas such as model risk, operational risk, and information security.

Governance oversight is owned by the Responsible AI Council (RAIC), a body co-led by the Compliance and the Chief Data Office. RAIC comprises senior members representing Risk, IT, Business Strategy and Group Internal Audit. RAIC meets every quarter and on a regular basis. Quarterly meetings are predominantly focused on AI Risk Framework topics. Regular meetings are predominantly focused on AI use cases, including AI Risk Assessments and approvals before going live into production.

RAIC owns and manages the AI Risk Management Framework (AI RMF). The AI RMF describes how Julius Baer addresses the risks associated with the development, deployment, and use of AI within the organisation. This framework adheres to JB Risk Management Framework and is complementary to existing group policies and guidelines.

RAIC governance applies to both third-party solutions as well as internally developed tools. All use cases, including Foundation Models, are reviewed by the Model Risk Management unit with a risk-based approach before deployment. The automation of controls, where feasible, is currently evaluated.

Implementation: The *General Chatbot*

One example of the governance and validation frameworks in action is the *General Chatbot*. The General Chatbot uses an open-weight model and operates without fine-tuning or RAG and serves as a general-purpose assistant rather than a domain-specific tool, to support users with tasks such as answering general questions, brainstorming, or drafting and paraphrasing texts.

It is hosted on a secure, on-premises environment which also supports fine-tuned or retrieval-augmented-generation (RAG)-based solutions, such as a tailored translation solution or the one that allows for querying investment research.

The model underlying the Chatbot has been updated and replaced since its initial launch several times already. This process been simple and smooth due to a pre-approval process for foundational models managed by the Model Risk Management function. This enables agile improvements without repeating full risk assessments for similar deployments. In addition, each use case is independently validated before deployment in a framework that utilises a risk-based assessment followed by a risk tiering of the use case.

The initial business case for launching the Chatbot was straightforward due to strong demand from employees and expected efficiency benefits. The tool did not involve high-risk or prohibited practices under the EU AI Act, allowing it to progress through the initial governance gates.

The primary risks identified were related to usage: potential reliance on inaccurate outputs (hallucinations), inappropriate prompts, or insufficient validation of results. These were addressed through two main measures:

1. **Training:** All users must complete a mandatory e-learning module covering responsible AI use and AI risks.
2. **In-tool guidance:** A visible disclaimer cautions users about the model's knowledge cutoff date, tendency to generate incorrect information, and the need to verify all outputs before use.

Deployment followed a controlled rollout. Access began with a cohort of approximately 500 employees who were trained through online sessions.

Once the mandatory e-learning had been introduced, user adoption soared: nearly every employee in Julius Baer has interacted with the Chatbot at least once, with a significant number in all functions and regions using it regularly.

Key Learnings

Several lessons have emerged from the implementation:

- Initially, Julius Baer was exploring use cases and their applicable governance practices. Now we see the need take a risk-based approach to ensure adequate governance and scalability.
- For generic solutions, like a General Chatbot, the main risk mitigation strategy is comprehensive user training.

By combining clear governance and user education, the General Chatbot demonstrates how financial institutions can deploy emerging technologies effectively and responsibly.

Prudential Implementation Example

Prudential Implementation Example

AI Governance in Practice: Lessons from Prudential's Implementation

Introduction

Prudential's approach to AI Governance is founded on the principles of ethical, safe and compliant use of AI. This framework is designed to build trust, ensure regulatory compliance, and support strategic objectives through the accelerated deployment of AI solutions aligned with Prudential's eight core AI ethics principles. Oversight is exercised by a multidisciplinary AI Governance Working Group (AIWG). The company's AI governance philosophy is aligned with its broader mission to be a trusted partner and protector for current and future generations, strengthening governance, transparency, and stakeholder confidence.

Enterprise AI Governance Landscape

Prudential has implemented a comprehensive framework for AI Governance, drawing on guidance from a broad range of domains including risk management, legal, compliance, privacy, security, and data science. The AIWG is a cross-functional taskforce that oversees the ethical, compliant, and strategic deployment of AI systems across the organisation. The AIWG includes senior leaders across key business and functional areas, ensuring that responsible, safe, and ethical AI is embedded across all AI initiatives. Organisational policies mandate rigorous, risk-based evaluation and ongoing compliance monitoring for all AI systems, whether developed internally or by third parties. This approach reinforces accountability and safeguards across the AI lifecycle.

Prudential's AI systems adhere to internal policies and standards that are applicable across all jurisdictions, including Singapore. All business units currently maintain their existing local governance while participating in global AI governance initiatives for AI-specific reviews hosted by Group AI Governance. Prudential regularly updates and assesses its risks, controls, and Key Risk Indicators (KRIs) to match evolving regulations and technology.

The company adapted the Model Risk framework to categorise AI use cases into risk tiers, managing them as part of its overall risk governance. This approach applies light-touch governance to low-risk AI systems and full governance requirements to higher-risk systems. Prudential's AI Registry encompasses the entire system lifecycle, from ideation and problem statement to post-deployment monitoring. Responsible AI knowledge is disseminated through training programs, internal guidelines, and a Group-wide Data and AI Community of Practice.

Prudential's technology infrastructure includes robust oversight and continuous monitoring, leveraging scalable technologies to enhance stakeholder value. The company is actively evolving its technology landscape, including enterprise platforms, to support responsible AI development and adoption and to enable automated ongoing monitoring.

Use Case Application: In-House Built PRUShield Chatbot

AI Use Case Description

The PRUShield Chatbot exemplifies strategic, high-impact AI innovation. Prudential's AI assistant is designed to address a critical business challenge: enabling Financial Representatives to confidently and quickly navigate complex queries across the PRUShield insurance suite with the support of a mainstream messaging application. By leveraging Gen AI, it delivers accurate, real-time answers from over 2,500 pages of product documentation, removing information barriers and empowering Financial Representatives to enhance client interactions and drive sales.

AI Use Case Ideation, Identification, and Onboarding

Once the project team identified the need to help Financial Representatives navigate thousands of pages of product details, they pitched the idea to management and got the green light. A sponsor, owner, and tech team were appointed to explore options; chiefly, whether to build in-house or to bring in a third-party solution. The team chose to build in-house, registered the use case in the AI Registry, and kicked off the internal development process.

Determine In-House Development or Third-Party Solution

If in-house development is selected, the project team would follow the AI governance steps embedded in the SDLC detailed below.

If a third party solution is selected (which does not apply to this case study), the project team registers the AI use case in the AI Registry and in parallel initiates the procurement process. This triggers Prudential's Third-Party Profiling and Risk Segmentation (TPPRS) intake, which drives third party profiling, risk segmentation, and domain reviews to surface AI specific risks early. The AIWG conducts preliminary AI due diligence at this stage. The depth of assessment scales with the third party risk materiality assessment. Some key TPPRS reviews includes:

- Data and Privacy – lawful purpose, data minimisation, retention, and transparency.
- Procurement – supplier due diligence, subcontractor visibility, and contractual obligations for AI change and incident reporting.
- AIWG – AI use case rationale, clarity, explainability, fairness and bias controls, robustness, and monitoring plans.
- Information Security – security posture and incident response readiness.
- Business Continuity Management (BCM) – resilience and recovery objectives.
- Operational Risk Management (ORM) – preparedness for operational disruptions, review of recent incidents and regulatory breaches, and insurance adequacy.
- Environmental, Social and Governance (ESG) – alignment with Prudential's ethical supplier practice standards.
- Risk and Compliance – applicable regulatory obligations and other risks.

A second full AIWG risk assessment and approval is required for all third-party AI systems prior to production, followed by periodic recertification based on the solution's risk tier.

Design Principles and Pre-emptive Risk Classification

Initial design concepts are reviewed by the Local Solution Review Committee (LSRC), the key governance checkpoint in Prudential's SDLC, to ensure that new solutions, enhancements, and architecture changes align with enterprise standards before implementation. Cross-functional teams present designs for a structured assessment of technical feasibility, compliance posture, and business impact, with deliberate, multi tiered stakeholder engagement. Adherence to Prudential's SDLC design principles is verified to ensure strategic alignment and compliance with internal and external requirements. The LSRC may require a return session before deployment to confirm execution against the approved design, reinforcing accountability and traceability across the lifecycle.

The LSRC recommended the PRUShield Chatbot to be endorsed for Group Solution Review Committee (GSRC) review since the messaging application was considered a non-standard application. The team presented to the GSRC and explained the rationale for using a mainstream messaging application as the front-end for the chatbot along with the guardrails implemented to minimise risk. GSRC's review comments and points for clarification were straightforward, and after each was responded to, the solution was approved by both the LSRC and GSRC.

Upon initial design approval, the project team registered the AI use case in the AI Registry and began early risk classification via AI use case materiality categorisation assessments. Systems are classified into risk tiers ranging from "Immaterial" to "Group Critical", which determines the depth of documentation and review required. The process centres on the AI Risk Assessment Questionnaire (AIRAQ), which evaluates systems across dimensions like transparency, fairness, privacy, and accountability. Higher-risk systems undergo formal review, including technical deep dives and AIWG voting. Performance metrics are assessed against ethical principles, and ongoing monitoring ensures continued compliance and alignment with internal policies and global standards.

The AIWG assigned a Moderate risk rating to the PRUShield Chatbot (given that it is used by Financial Representatives to support customer engagement) and mandated annual recertification to ensure ongoing fit for purpose.

Build, Customisation, and Testing

The chatbot uses retrieval augmented generation (RAG), answering only from publicly available, approved product sources and FAQs and refusing out of scope or account specific queries. Data was sourced from Product Information Packs, publicly available information about PRUShield/PruExtra (including Premium Tables) and PruPanelConnect found in the corporate website, and 500+ FAQs provided by Subject Matter Experts (SMEs). These are corporate product documentation of high quality and publicly available materials that are already being shared with Financial Representatives. Due to this information scope, privacy issues were assessed as non-material. Mitigation steps included implementing guardrails such as disclaimers and reminders to always validate the chatbot's response before sharing with customers. Retrieval settings and mandatory citations are versioned with the index for transparent, auditable, secure by design responses and grounding aware evaluation. User Acceptance Testing (UAT) with business users and SMEs validated end to end behaviour (representative questions, ambiguity handling, refusals) and the system tracks Faithfulness and Context Precision/Recall as go live gates and regression guards, reported during UAT and via scheduled sampling.

The UAT testing process took approximately 4 months. The process involved a review of chatbot responses from six SMEs from the Product, Pricing, Operations and Health teams. The initial accuracy rate was 41% pass, based on the assessment of responses by the SMEs. Business owners and SMEs expected highly accurate responses aligned with their domain expertise with accuracy and consistency >95%. However, the chatbot's knowledge base primarily derived from product documents often lacked sufficient depth or clarity in certain areas.

This led to dissatisfaction when responses appeared vague or incomplete. To address these challenges, questions were categorised into two buckets:

- Bucket A: The chatbot's response must always be correct and consistent.
- Bucket B: The chatbot can be creative but still factual.

SME reviewers adopted a more consistent evaluation approach during the review of over 500 chatbot questions and responses. Through this iterative process, the chatbot progressively improved and achieved 99% accuracy, meeting the defined testing criteria.

Prudential opts out of provider model training by configuration, opting instead for an inference only, zero/ controlled retention instance. Guardrails enforce publicly available, approved product sources and FAQs only, with no personalised advice/decisioning permitted, and also require mandatory citations. Centralised AI-specific infrastructure such as an adapter layer applies authentication and parameter validation, Personally Identifiable Information (PII) detection/masking at ingress, toxicity/jailbreak screening, and unsafe format filtering. Inputs and uploads are screened with a redaction pipeline before processing; data loss prevention applies to prompts and outputs; egress is restricted to approved channels or destination to minimise leakage and ensure compliance; and all prompts are logged for investigations and governance.

Pre-deployment Reviews & Sign Offs

Before deployment, the PRUShield Chatbot underwent rigorous testing and governance to ensure readiness across performance, safety, and business alignment. Engineering validated rejection handling, while UAT with business users and SMEs confirmed end to end behaviour, performance was tracked on dashboards, and security controls such as guardrails, jailbreak rejection, toxicity filtering, and PII masking met internal standards. UAT started at the end of Feb 2025 and lasted until end of May 2025. UAT signed off was obtained in the first week of June 2025.

Governance for the use case includes the AIWG, which served as a multi-domain lead reviewer (ethics, safety, performance, privacy, legal, Model Development Report, etc.), Risk & Compliance (alignment with internal and external requirements), and LSRC (architecture, data flows, and controls), with input from security, engineering, and architecture teams.

The Change Advisory Board (CAB) granted final approval after all parties upstream had signed off. All necessary artefacts – such as test results, guardrail outcomes, and architecture documents – were archived in the governance repository. Deployment followed a phased rollout with a checklist and rollback plans; users would also undergo training to ensure effective and responsible usage.

Post Deployment Monitoring and Recertification

Following deployment, the PRUShield Chatbot is continuously monitored and formally recertified to keep performance, safety, and compliance aligned with internal and external governance expectations. Runtime monitoring tracks failures/incidents, safety filter activations, and retrieval health (stale/missing product content), with all indicators centrally logged for traceability and review. Incident management follows enterprise protocols; if operations are materially impacted, contingency plans are in place to remove or replace the affected AI component, especially for high risk use cases to preserve business continuity and governance integrity.

Introducing AI into existing non AI solutions triggers a fresh AIWG assessment to uphold Prudential's AI ethics principles. All AI systems, whether in house or third party, require periodic AIWG recertification, covering operational/quality metrics, guardrail revalidation, risk tier reassessment, and confirmation of continued ethical compliance/approval status. The cadence of recertification is set by risk materiality tier (higher-risk requiring more frequent recertification), and all reviews/outcomes are recorded in the AI Registry.

The AI project team runs sampled evaluations using Faithfulness and Context Precision/Recall; regression gates trigger governance review when thresholds are breached. A user feedback loop captures flagged responses and satisfaction scores for triage and retraining. Toolkits including Fairlearn (Traditional AI), LLM as a Judge, AI Verify Moonshot, and G Eval (Gen AI) are available to support these efforts.

Automation of Governance Process

Multiple platforms were considered and utilised throughout the governance process. Due to the complexity and critical nature of technology governance, most steps required direct involvement and oversight by SMEs; those tools are primarily supported coordination and documentation. Prudential is constantly reviewing platforms and tools to enhance and improve the governance process.

AI Governance Implementation Challenges & Lessons Learnt

As AI governance continues to evolve, several lessons have emerged that offer valuable insights for strengthening oversight practices. While some of these challenges have already been addressed through targeted improvements, others continue to highlight opportunities to clarify stakeholder expectations, refine assessment processes, and improve coordination. Together, these reflections provide a foundation for uplifting governance maturity, enabling more consistent, scalable, and transparent review mechanisms.

Aligning Expectations: During UAT, some stakeholders expected highly detailed, domain-specific answers. As the chatbot relies on approved product documents, content gaps sometimes produced general responses, highlighting the need for clearer scope and stronger content curation.

Refining Risk Assessments: Differences in interpreting certain risk assessment questions caused delays. Clearer guidance and sharper assessment processes would improve consistency.

Reducing Overlap: Similar questions surfaced across multiple forums and reviews. Coordinating checkpoints and eliminating duplication can improve efficiency without reducing rigour.

Rating Risk of AI Systems: Identifying and rating the risks of an AI use case in early phases is difficult, especially as many critical details are not known. To resolve this issue, Prudential has adopted a 2-phased risk rating methodology. In Phase 1, an initial risk rating is provided based on factors such as the business process in which AI is used, the user's maturity, the AI capability's maturity, etc. A final risk rating is provided in Phase 2 after the AI assessment is conducted by a multidisciplinary AI governance body.

Limited Expertise and Resources: Scaling AI governance is challenging due to insufficient domain and local expertise and constrained resources, especially as AI use cases grow rapidly in volume and complexity.

Strengthening AI Literacy: While internal courses have reached a broad audience, there remains room to deepen the understanding of AI across roles and functions within each domain.

Ongoing Maintenance: After go-live, the main challenge was sustaining adoption and keeping the chatbot's knowledge current. Frequent changes to panel clinics, doctors, and upcoming product updates meant that parts of content quickly became outdated. To address this, we established a regular process for updating and testing the chatbot's knowledge base, alongside ongoing performance monitoring. This ensures accuracy and relevance for users. The experience highlighted that post-launch success depends on proactive content management and continuous evaluation, not just initial deployment.

Future Plans and Strategies

Evaluation toolkits such as Fairlearn (for Traditional AI), LLM as a Judge, AI Verify Moonshot, and G Eval (for Gen AI) are currently used manually. Group AI is working to integrate these tools into Machine Learning Operations (MLOps) and Large Language Model Operations (LLMOps) pipelines to automate statistical evaluations for high risk systems. In parallel, Prudential is exploring additional tools and platforms to streamline governance processes, reduce manual effort, and improve efficiency across the AI lifecycle.

The AIWG is implementing a hub-and-spoke model to govern AI across the Group and Local Business Units (LBUs), balancing central oversight with local adaptability. In this model, the Group AIWG (the "hub") provides standardised frameworks, the AI governance platform, risk rating tool, and expert guidance, while LBUs (the "spokes") handle local implementation, initial reviews, and context-specific adaptations. This federated approach enables scalable, consistent, and efficient AI governance – allowing LBUs autonomy for low-risk AI systems while ensuring that moderate- and higher-risk initiatives receive rigorous controls and Group oversight. Automation through the AI governance platform streamlines intake, risk assessment, and monitoring, while embedded compliance checks at both local and Group levels ensure ongoing quality and ethical standards throughout the AI lifecycle.

Prudential is exploring ways to ease resource constraints in AI governance, including the potential use of external contractors to support review activities. Efforts are also underway to strengthen internal expertise through targeted training and knowledge-sharing initiatives.

Investment Firm Implementation Example

Investment Firm Implementation Example

Enterprise AI Governance Landscape

An Investment Firm has established an AI Governance Ecosystem for the Responsible Use of AI in the organisation. This ecosystem is organised with clear roles and responsibilities covering enterprise adoption, policy formation, information security, and engineering standards.

It consists of three bodies:

AI Council

- Directs the Investments Firm's overall development and implementation of AI.
- This is a business use case-driven body with representation from senior leadership, as well as subject matter experts from the Technology, Data, and Investment Divisions.

AI Governance Workgroup

This Workgroup focuses on three key areas:

- **Legal and Regulatory Compliance:** Continuously tracks, assesses, and advises on the implication of AI laws and regulations to the organisation.
- **Responsible Business Usage:** Established the Enterprise AI Policy and Ethical Guardrails to guide and govern all AI business usage across the organisation.
- **Cybersecurity & Info-security:** Implements protection measures for AI and enterprise systems to guard against cyber threats and ensure reliable operations.

This Workgroup has representation from control functions, including Legal and Compliance, Cybersecurity, and Governance and Client Relations.

AI Engineering Workgroup

This group is comprised of subject matter experts from the Technology and Data departments and focuses on AI engineering governance. They are responsible for defining a standardised AI/ML Ops architecture, engineering standards, and business templates and best practices to drive and govern enterprise implementation of AI tools and platforms.

AI Governance Operationalisation

AI workloads typically fall into three categories: Software as a Service (SaaS), Commercial Off-The-Shelf (COTS), and internally developed solutions. SaaS-based AI solutions tend to pose less risk, particularly when sensitive data is not involved. A risk-based approach enables the Investment Firm to concentrate its risk management efforts where they are most critical, while supporting the accelerated and secure adoption of AI technologies.

To empower business units to harness the benefits of AI efficiently, the IT Risk Management team has developed an assessment method tailored for different tires of risks which AI systems present.

The Investment Firm's approach to developing AI governance-related skills and knowledge is to hire AI talent and upskill the existing staff.

Key documents include the AI Policy and the IT and IT Security Management policies and standards. The Technology Risk team/ITRM manages AI risks as part of their AI risk assessments. All SaaS AI products go through the AI Risk Assessment, Technology Risk Assessment, and Third-Party Risk Assessment.

High-Risk Use Cases

High-risk use cases are those that meet any of the following criteria:

- Processing sensitive data.
- Data is preserved for training/fine tuning.
- Automated actions (no human in the loop).

Use Case Application

One AI use case implemented in the institution is around corporate policies search and answer. The goal of the project is to allow employees to quickly find and retrieve answers to policy-related questions using AI.

The use case was presented at the AI Council and the approach to development and implementation was approved.

A third-party solution was identified for the search and answer, and went through:

- AI specific cybersecurity review to ensure that both AI- and technology-related risks were identified and mitigated.
- Legal and regulatory compliance to ensure that the AI solution and output were in line with regulatory requirements.
- Engineering implementation reviews to ensure that data flow to AI LLMs had the appropriate guardrails and security hardening.

Implementation Challenges

Key challenges that the firm has encountered in the implementation of AI governance include:

- Global regulatory variability – like the EU AI Act versus more fragmented state-level AI policies in the US – can lead to implementation complexity. The Investment Firm must navigate inconsistencies in legal definitions, data standards, and regulatory benchmarks.
- Robust governance frameworks can sometimes stifle innovation, especially when they adopt a risk-averse posture.
- Incomplete or biased datasets can lead to algorithmic biases that may harm certain user groups.
- Interpreting the predictions of high-complexity AI models ("black-box" systems) is difficult, making it hard to ensure accountability.
- A lack of comprehensive AI literacy among business units causes inconsistencies in their ethical decision-making and risk management.

Lesson Learned

Common issues observed with AI products include:

1. Inadequate protection against adversarial manipulation.
2. Insecure AI supply chain practices.
3. Observability as an afterthought.
4. Lack of reference architecture design best practices (SaaS, Agentic AI, etc.)