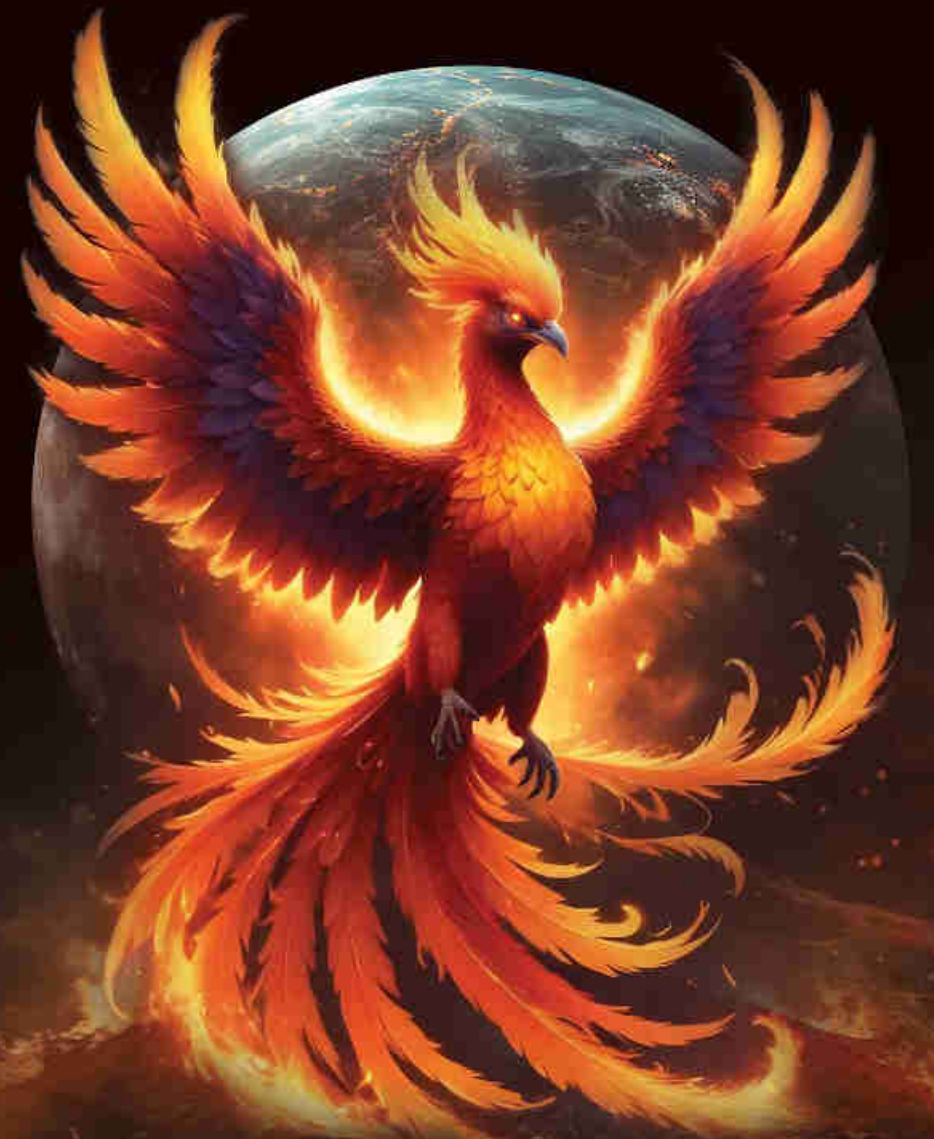


iERP®
Institute of Enterprise Risk Practitioners

GLOBAL CONFERENCE 2025

ENTERPRISE RISK MANAGEMENT:
RISING FROM THE ASHES

GEN AI

AND

MODEL RISK MANAGEMENT

ROCHAK AGRAWAL

EXECUTIVE DIRECTOR AND HEAD OF RISK MANAGEMENT

UBS TECHNOLOGY

About the Speaker – Rochak Agrawal

Institute of Enterprise Risk Practitioners

Executive Director, UBS (Singapore)
Head of Technology Risk – Investment Bank Operations Technology
Head of Agentic AI Delivery – Wealth Management AI Technology
18+ years of experience in banking technology, risk, and AI innovation.
London | Hong Kong | Singapore
Board Member – Singapore International Chamber of Commerce & Hong Kong University of Science and Technology (HKUST)
Intl. Speaker – Dubai | Kuala Lumpur | Singapore | Hong Kong | Seoul
Master's degrees from London and Columbia Business School
Computer Engineer NTU, Singapore(First Class Honors) and Penn State, USA (Dean's List)

GLOBAL CONFERENCE 2025
ENTERPRISE RISK MANAGEMENT: RISING FROM THE ASHES

Agenda for today (45 mins)

- **Why This Matters Now -** The shift from GenAI pilots to enterprise adoption — balancing speed, value, and control.
- **Understanding Model Risk in GenAI -** Failure modes unique to GenAI, and why context amplifies risk.
- **Synthetic Data: Benefits, Uses & Testing -** What it is, practical examples, and how to test for utility, privacy, and bias.
- **Build vs Buy (and Blend) Decision Framework -** Criteria, practical examples, and controls to guide the right choice.
- **The Commoditization Curve -** Where the real value moves as base models become utilities.
- **Operating Model & Controls -** Policies, governance, safety gates, and monitoring for evidence by design.
- **90-Day Action Plan** - A pragmatic, phased roadmap to go from concept to controlled scale.
- **Interactive Scenarios [Audience interactive] -** Live Build/Buy/Blend exercises to apply the framework.

Why Now: From Pilots to Production

- Embedded already: Service Desk , dev tools, research, reporting, controls testing
- Two pressures: Rapid ROI (from board) and audit-ready evidence of control (from regulator)
- Goal: Measurable value and measurable safety
- Design for scale: Scope, Success, HITL, testing, monitoring from Day 0
- Analogy: Building a highway

“Models are commoditizing; governed integration is the difference”

Model Risk in GenAI

- Model risk: Harmful/incorrect/non-compliant outcomes from AI
- GenAI failure modes: Hallucination, privacy leakage, jailbreaks, bias/toxicity
- Risk amplifiers: Long context, Long memory ,tool use, autonomy, feedback loops
- Context matters:
  - Separate “cause” from “context”
  - Safe in sandbox ≠ safe in production
- Analogy: Very Fast, Very Helpful intern – you still review outputs!

Synthetic Data 101: What, Why, Types

- Definition: Artificially generated data that mimics real distributions
- Why use it: Privacy-by-design, rare-case coverage, balancing speed
- Common types: Tabular/time-series, text/dialogue, images/docs, logs
- Key caveat: Fidelity ≠ truth; bias can persist or amplify.
- Analogy: It's a shield not a cloak

Model Risk in GenAI

Real (c

Cus
Age

Synthe

Cus
Age

- Sam
- Keep

Real (energy usage):

[redacted]

```
2024-07-01  10.5 kWh
2024-07-02  12.1 kWh
2024-07-03   9.8 kWh
```

Synthetic:

[redacted]

```
2024-07-01  10.4 kWh
2024-07-02  12.2 kWh
2024-07-03   9.9 kWh
```

Synthetic Data: Risks & Utility Tests

Risks:

- Privacy risk: Memorisation of real records.
- Bias preservation/amplification: Skewed patterns remain or worsen.
- Low fidelity: Unrealistic patterns that don't match production.
- Utility overestimation: Adds noise, lowering real-world performance.

Utility & Safety Tests:

- Downstream performance comparison: Real vs Real+Synthetic. Calculate F1 score.
- Rare-class coverage: Match or slightly oversample important low-frequency cases.
- Drift & robustness tests: Check stability under changing patterns.
- Membership inference: Detect leakage of real records.
- Nearest-neighbor distance: Ensure synthetic ≠ direct copy of real records.

Analogy: think of synthetic data as a shield, not a cloak

Recap Template: Synthetic Data Test Plan

1. Objective & scope of synthetic dataset
2. Utility metrics: downstream accuracy, rare-class lift
3. Privacy tests: membership inference, NN distance
4. Bias/fairness evaluation plan
5. Governance artifacts: dataset card, lineage, risk scores
6. Acceptance criteria & sign-off

GenAI Models: The Commoditization Curve

- Base models → utility: Rapid releases, lower cost
- Value shifts up-stack: Data, retrieval, guardrails, workflows
- Design for swap-ability: Orchestration layers
- Winning pattern: **Thin customization + thick governance + deep integration**
- **Engine** -> Base Model
  Fuel -> Your data
  Brakes and Seatbelts -> Your guardrails
  Driver Training -> Your operating model

Internal vs Third-Party Models: Clear Trade-offs

- **Internal/open-weight**: Pros (control, transparency, cost), Cons (ops burden, talent). Think: Training your own models using ML studio (not foundation models)
- **Hosted/API:** Pros (quality, speed, safety), Cons (lock-in, data risk, explainability). Think: Azure hosted OpenAI models, License, Vendor solutions
- **Middle Hybrid path:** RAG + adapters, secure gateways, contractual clauses “No train”. Think: hosted models but internal RAG layers (embeddings, data filters).

Build vs Buy Decision Matrix

- Criteria: Strategic fit, Data advantage, Risk constraints, Time/talent, TCO, Vendor terms
  - **Is it core IP/differentiating?** If yes → bias to **Build/Adapt**.
  - **Do we have a data advantage?** If yes → **Build/Adapt** to exploit it.
  - **Are there hard constraints?** Residency, explainability → **Self-host/Open-weight**.
  - **Do we need results in weeks?** If yes → **Buy** to start; design portability.
  - **What is true TCO?** Include evals, safety, monitoring, updates, staffing, **switching**.
  - **Can we exit?** Demand **no-train**, deletion SLAs, export, pricing caps/ramps.
- Scoring: 1–5 scale for each criterion
- Weighting: Assign importance per org strategy
- Total score: Guide decision (Build / Buy / Blend)

Operate with Evidence: Controls & De-Risk

- **Policies**: AI use, model risk, data handling, HITL
- **Governance**: Model Inventory, Risk tiering, approvals, validation (second line)
- **Safety**: Evaluation suites, jailbreak tests, red-team logs
- **Monitoring**: Logs, alerts, incident mgmt
- **Operating Model**: AI Steering Committee, Model Risk Committee, Deliver Pods, RACI
- **Controls**: Contractual (no-train on prompts/outputs), deletion SLAs, architectural, assurance, monitoring, change mgmt

How to start: 90-Day Plan

- **90-day plan:**
  - 0–30 days pick use cases (high –ROI) and establish policy baselines
  - 31–60 days build RAG pipeline, implement safety gates, pilot with HIL
  - 61–90 days scale & validate, negotiate no-train contracts, setup independent validation

AUDIENCE Turn

GLOBAL CONFERENCE 2025
ENTERPRISE RISK MANAGEMENT:
RISING FROM THE ASHES

Scenario 1: Summarising 2M Historic Customer Emails

- Goal: *"We need an AI to summarise 2 million historic customer emails for training a support bot. Build, Buy, or Blend?"*
- Options: **Build, Buy, or Blend?**
- Consider privacy, speed, and compliance

Answer: **Blend**

- Hosted LLM for language summarisation quality and speed
- PII scrubbing before sending data outside
- In-house RAG layer retrieves only masked, relevant excerpts [e.g next slide]
- Meets privacy & compliance while delivering quickly

Masked Email → Embedding → Vector DB

- Original email: *Dear support, my account 123456 was double charged for product X on Jan 5.*
- **Step 1:** Masked email: Dear support, my account [ACCOUNT_NUMBER] was double charged for product X on Jan 5.
- **Step 2:** Generate embedding (vector representation of meaning) using open-weight model (e.g., SentenceTransformers).
  - Example vector (truncated): [0.021, -0.143, 0.532, 0.287, ...] for Step 1
- **Step 3 :** Store embedding + metadata (date, product, issue type, masked text) in secure vector DB (e.g., Milvus, Weaviate, Pinecone).

Retrieval → Prompt → Safe Summary

- **Step 1:** Query: *'All refund-related complaints in January 2023'*
- **Step 2:** Vector DB converts query to embedding and retrieves most similar masked chunks.
  - Retrieved examples:
  - 1. Dear support, my account [ACCOUNT_NUMBER] was double charged for product X on Jan 5.
  - 2. I was charged twice for my subscription in Jan.
- **Step 3:** Prompt to hosted LLM: Summarise key themes; exclude personal identifiers.
- **Step 4:** LLM output: - Multiple customers experienced double charges in early January. - Issues affect both purchases and subscriptions. - Customers request immediate refunds.

Scenario 2: Fraud Detection Synthetic Data

- Goal: *“We need a GenAI to generate fake transaction histories for fraud detection model training. Build, Buy, or Blend?”*
- Options: **Build, Buy, or Blend?**
- Consider domain specificity and privacy

Answer: **Build**

- Fraud patterns are highly domain-specific
- Full control over generation rules and validation
- Run privacy tests: membership inference, re-identification
- Hosted options may lack transparency for sensitive data

Scenario 3: Developer Copilot

- Goal: *“We want a coding assistant for internal devs. Build, Buy, or Blend?”*
- Options: **Build, Buy, or Blend?**
- Consider speed, quality, and IP protection

Answer: **Buy first**

- Hosted LLMs trained on massive codebases deliver quality fast
- Add repo scoping and secrets filters
- Measure test pass-rate & defect reduction
- Evaluate open-weight build later for cost and IP control

Scenario 4: **Customer Complaint Classification**

- Goal: *“We need to classify incoming customer complaints into risk categories in real time. Build, Buy, or Blend?”*
- Options: **Build, Buy, or Blend?**
- Consider internal data sensitivity and user experience

Answer: **Build**

- Classification can be a smaller model fine-tuned on your proprietary labelled data. lightweight to run in-house with minimal latency.
- Gives full control and explainability for compliance/audit (important in risk categorisation).
- Hosted may be overkill and risk exposing sensitive complaint content.

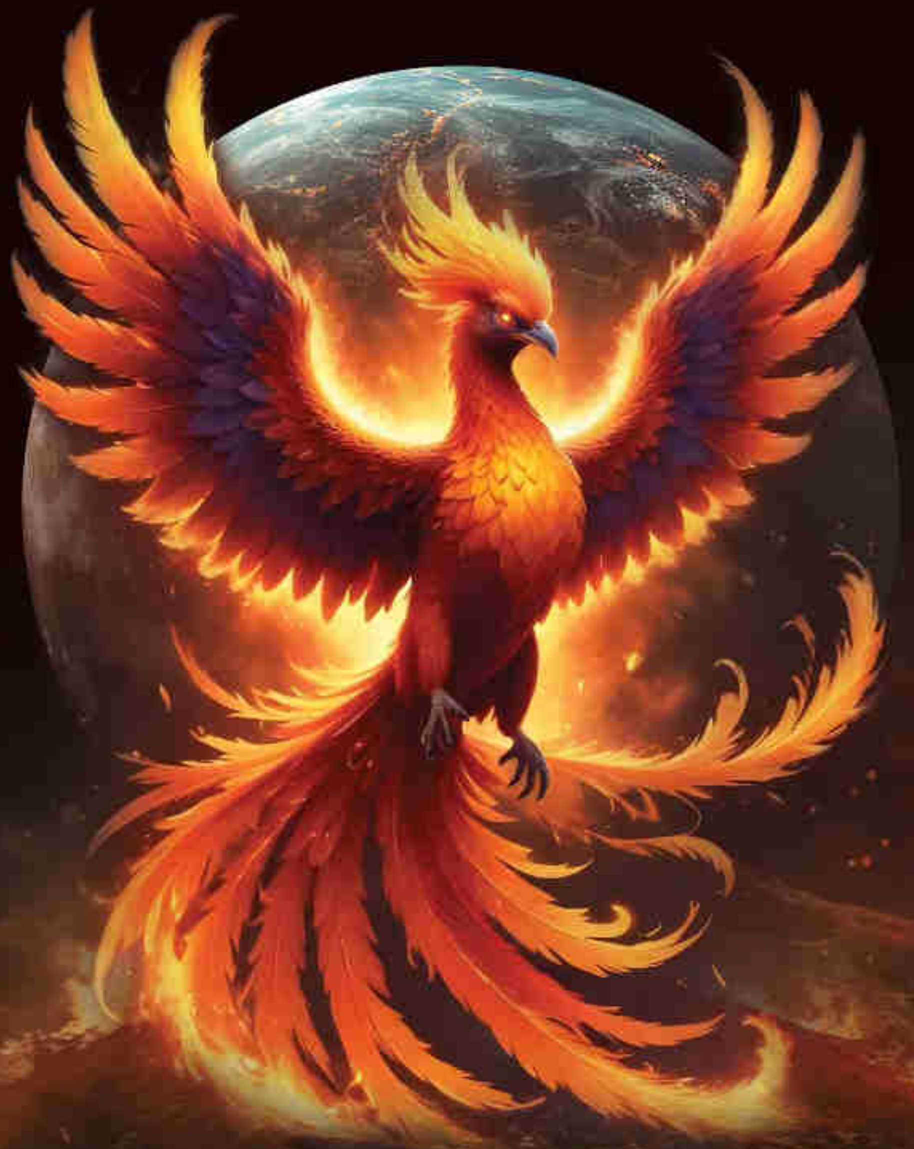
Scenario 5: HR Policy Q&A Bot

- Goal: *"We need a chatbot for employees to query HR policies, benefits, and procedures. Build, Buy, or Blend?"*
- Options: **Build, Buy, or Blend?**
- Consider internal data sensitivity and user experience

Answer: **Blend**

- Hosted LLM for natural conversation
- In-house RAG for secure HR document retrieval
- Only safe excerpts sent externally
- Delivers quality while safeguarding sensitive info

iERP®
Institute of Enterprise Risk Practitioners



GLOBAL CONFERENCE 2025

ENTERPRISE RISK MANAGEMENT: RISING FROM THE ASHES

THANK YOU

InstituteOfEnterpriseRiskPractitioners

institute-of-enterprise-risk-practitioners

ierp_official