The AI Consultant Career Roadmap

From Foundational Skills to AI Practice Leader

A step-by-step guide to building a career in AI consulting with self-assessment tools to plan your own path.

1. Introduction

Generative AI is no longer confined to pilot projects and innovation labs. Banks are actively using it for customer service, document processing, fraud operations, software development, research, compliance support, and internal knowledge management. As adoption accelerates, the question practitioners face is no longer whether to use generative AI, but how to govern it responsibly.

Traditional model risk management (MRM) was built around models with defined mathematical structures and relatively stable behavior. Generative AI does not always fit that mold. Outputs are probabilistic, models can hallucinate, behavior can shift as underlying systems are updated, and many banks rely on third-party foundation models with limited visibility into how they were built or trained.

This guide exists to close that gap. It translates established MRM principles into a framework specifically shaped for generative and agentic AI, so that governance keeps pace with how the technology actually behaves — not just how legacy models used to behave.

Why GenAI Needs a Distinct Governance Layer

Three characteristics set generative AI apart from the models most MRM programs were designed for:

- Non-determinism — the same prompt can produce different outputs across runs, which complicates traditional validation and reproducibility expectations.

- Opacity — many banks license foundation models rather than build them, which limits visibility into training data, architecture, and update cycles.
- Dynamic behavior — a vendor-side model update can silently change how a deployed application performs, even when nothing in the bank's own environment has changed.

Example — Non-determinism in practice: *A bank's internal research assistant, given the same question about a regulatory threshold on two different days, returns a slightly different figure each time. Neither answer is obviously wrong, which makes the discrepancy harder to catch than a traditional model bug.*

Where This Framework Fits Relative to SR 26-2

In April 2026, the Federal Reserve, OCC, and FDIC issued SR 26-2, revised interagency model risk management guidance that superseded SR 11-7. The new guidance takes a more risk-based approach and explicitly places generative and agentic AI models outside its specific scope — while still expecting banks to apply appropriate governance and controls to technologies it does not directly cover.

In practical terms, that means SR 26-2 will not tell you exactly how to govern a GenAI chatbot or an agentic workflow. This framework is designed to fill that space: it borrows the risk-based spirit of SR 26-2 and applies it specifically to the characteristics of generative AI.

Who This Guide Is For

- Model risk officers and validators evaluating GenAI systems for the first time.

- Compliance and legal teams assessing regulatory exposure of GenAI use cases.
- GenAI product owners and engineering leads who need to build governance into a system before launch, not retrofit it afterward.
- Internal audit teams testing whether GenAI governance claims match actual practice.

2. Regulatory Context at a Glance

Understanding where GenAI sits relative to existing regulatory guidance is the starting point for any governance conversation. The table below summarizes the shift from SR 11-7 to SR 26-2.

	SR 11-7 (2011)	**SR 26-2 (April 2026)**
Approach	Prescriptive, uniform expectations across model types	Risk-based, tailored to institution size and complexity
GenAI / agentic AI	Not addressed (predates the technology)	Explicitly outside specific scope, due to rapid evolution
Governance expectation	Applied uniformly via model validation lifecycle	Banks must still apply 'appropriate' governance to tools not covered
Status	Superseded	Current guidance

What's Explicitly Out of Scope — and Why That's Not a Free Pass

SR 26-2's decision to exclude generative and agentic AI from its specific scope reflects how quickly the technology is changing, not a judgment that it is low-risk. The guidance is clear that banks are still expected to apply appropriate risk-management and governance practices to any tool or system it doesn't directly cover. In effect, regulators are asking institutions to build their own risk-based governance — which is exactly what this framework provides structure for.

Key Regulatory and Standard-Setting Bodies Referenced

- Federal Reserve, OCC, FDIC — joint issuers of SR 26-2 and the primary U.S. banking supervisors for model risk.
- Bank for International Settlements (BIS) — has flagged hallucinations, inconsistent outputs, data confidentiality, cybersecurity, third-party dependencies, and explainability as core GenAI risks in finance.
- Financial Stability Board (FSB) — has identified third-party dependencies, cyber risk, model risk, and governance gaps as AI-related vulnerabilities at a systemic level.

Example — Applying this section: *A bank preparing for an examination cites SR 26-2 as its baseline MRM standard, then documents a supplementary internal policy — modeled on this framework — specifically for GenAI and agentic systems, explicitly noting that it fills the gap SR 26-2 leaves open.*

3. The Five-Step Framework

Governance should start with the use case, not the underlying technology. The same foundation model can power a low-risk internal writing tool and a high-risk credit-decisioning assistant — the controls required depend entirely on how it's used. The five steps below are designed to be worked through in order, with each step narrowing the scope of what the next step needs to cover.

Step 1: Define the Intended Use

Before anything else, document exactly what the system is supposed to do, for whom, and where its boundaries lie. This becomes the reference point every later evaluation is measured against — including whether the system is later used outside its intended purpose, which is itself a common source of model risk.

Worksheet prompts to complete for each use case:

- Purpose — What specific task or decision does this system support?
- Users — Who interacts with it directly (employees, customers, both)?
- Boundaries — What should the system explicitly not be used for?
- Output — Does the system produce text, a decision, a recommendation, or an action?

Example — Step 1: *A bank defines the intended use of a new tool as: 'Summarize customer complaint emails into a structured ticket for the operations team. Not intended to draft customer-*

facing responses or make resolution decisions.' That boundary alone determines much of the risk tier in Step 2.

Step 2: Classify the Risk

Not every GenAI application warrants the same level of scrutiny. A basic internal writing assistant needs lighter governance than a system that influences lending, fraud, or customer outcomes. Classifying risk lets an organization allocate validation effort, monitoring, and human oversight where they matter most, rather than applying uniform — and often disproportionate — controls everywhere.

A simple three-tier structure works for most institutions:

Tier	Characteristics	Example Use Case
Low	Internal only, no customer impact, human reviews all output before use	Drafting internal meeting summaries
Medium	Customer-adjacent or influences internal decisions, but a human remains the final decision-maker	Drafting (not sending) customer email responses
High	Directly influences lending, fraud, investment, or customer-facing decisions with limited human review	Flagging transactions as likely fraudulent for auto-hold

Step 3: Evaluate the Data

Data governance is foundational to GenAI risk management, because the data a system can access often determines its worst-case failure mode. Practitioners should be able to answer each of the following for every use case:

- What information enters the system — structured data, free text, documents, or a mix?
- Does it contain confidential or customer information, and if so, what category (PII, account data, credit information)?
- Where is the data stored, and for how long?
- Who — and what systems — can access it, including the model provider itself?
- Can prompts or outputs be retained, logged, or used for further model training by a vendor?
- How is sensitive information protected in transit and at rest?

The BIS has specifically flagged data confidentiality and data-management practices as core concerns when financial institutions adopt GenAI — this step operationalizes that concern into a checklist a team can actually complete.

> ***Example — Step 3:*** *Before connecting a GenAI tool to a document repository, a bank's data team discovers the repository includes unredacted customer statements. The use case is paused until a redaction or access-scoping layer is added — a catch that only happens if this step is done before deployment, not after.*

Step 4: Test the Model and Application

Testing a GenAI system means more than confirming it produces plausible-sounding answers. It requires evaluating the system across several dimensions, ideally under conditions that resemble its actual production workflow rather than a clean lab environment.

- Accuracy — Are factual outputs correct against a known-answer test set?
- Consistency — Do repeated runs of the same prompt produce materially different results?
- Bias — Does the system treat comparable inputs differently based on protected characteristics or proxies for them?
- Hallucination rate — How often does the system state something confidently that isn't true, and under what conditions is this most likely?
- Robustness — How does the system behave with ambiguous, adversarial, or out-of-scope prompts?
- Security — Can the system be manipulated through prompt injection or data exfiltration attempts?

Testing should reflect the real workflow: a model that performs well in isolated testing can still fail once embedded in a broader application with retrieval steps, tool calls, or agentic actions layered on top.

> ***Example — Step 4:*** *A fraud-flagging assistant passes accuracy testing on a static dataset but, when tested against live-formatted transaction feeds with occasional missing fields, begins hallucinating fraud indicators. The gap is only caught because testing used production-like data rather than a curated sample.*

Step 5: Establish Human Oversight

Human oversight should be scaled to the risk tier established in Step 2, not applied uniformly. For high-impact decisions, meaningful human review is essential; for lower-risk tasks, automated workflows with monitoring and escalation paths may be sufficient.

- High tier — Mandatory human review and sign-off before any output is acted on.
- Medium tier — Spot-check review with clear escalation triggers for anomalous outputs.
- Low tier — Automated monitoring with periodic sampling; no mandatory pre-review.

The critical design principle: 'human in the loop' is not a checkbox. The person reviewing an output needs enough context, authority, and time to genuinely identify and challenge an incorrect result — not simply click 'approve' under time pressure.

> ***Example — Step 5:*** *A bank initially routes all GenAI-drafted credit memos through a single reviewer who has ten seconds per memo due to volume. After an internal audit flags this as ineffective oversight, the bank redesigns the workflow so the system pre-flags low-confidence sections for closer review, giving the human reviewer's attention where it's actually needed.*

4. Key Risk Categories Reference Sheet

The five risk categories below recur across nearly every GenAI use case in banking. Each entry includes a working definition, why it matters specifically in a banking context, and one concrete mitigation to start with.

Hallucination & Accuracy

The system generates plausible but false information.

Why it matters in banking: Outsized in banking because outputs may relate to regulations, customer data, or financial analysis where errors carry real consequences.

Start with: Require citation or source-grounding for factual claims; flag low-confidence outputs for review.

Explainability

Difficulty describing how an AI-supported outcome was produced.

Why it matters in banking: Banks may need to explain an outcome to a regulator or customer; proprietary foundation models limit visibility into how a result was generated.

Start with: Maintain decision logs capturing inputs, key intermediate outputs, and the human review step.

Bias & Fairness

The system reproduces or amplifies biases present in training data or design.

Why it matters in banking: Especially critical in customer-facing or decision-support processes where disparate treatment carries legal and reputational risk.

Start with: Test outputs across demographic proxies before and after deployment; monitor for drift over time.

Data & Privacy

Sensitive customer or proprietary information is exposed, retained, or misused.

Why it matters in banking: Customer trust and regulatory compliance depend on knowing exactly what data an AI system can see and retain.

Start with: Scope data access tightly; confirm vendor data-retention and training-use policies in writing.

Third-Party / Vendor Risk

Dependence on external foundation-model providers, cloud platforms, and APIs.

Why it matters in banking: Raises questions of service availability, model-version changes, security, concentration risk, and vendor transparency.

Start with: Maintain a vendor risk register; require advance notice of material model changes where contractually possible.

5. Risk Classification Matrix (Tool)

Use this table to classify each GenAI use case in your inventory. Fill in one row per use case; the risk tier determines the controls and oversight level required from Step 5.

Use Case	Data Sensitivity	Decision Impact	Risk Tier	Required Controls
Meeting notes summarizer	Low	None	Low	Periodic sampling
Draft customer email responses	Medium	Indirect	Medium	Spot-check review
Fraud transaction flagging	High	Direct	High	Mandatory review + sign-off
[Your use case]				
[Your use case]				

Tip: *revisit this matrix whenever a use case's data access, user base, or decision authority changes — risk tier is not a one-time label.*

6. Practitioner's Checklist

Work through this checklist before deploying any GenAI system. Each item includes space to record an owner and date, so the checklist doubles as an audit trail.

Item	Question	Owner	Date
Purpose	What problem is the system solving?		
Impact	What happens if the output is wrong?		
Data	What information does the system access?		

Model	Which foundation model or AI technology is being used?		
Vendor	Who provides and maintains it?		
Testing	How has performance been evaluated?		
Security	What prevents unauthorized access or misuse?		
Monitoring	How will changes and failures be detected?		
Human oversight	When must a person review or approve an output?		
Documentation	Can the organization demonstrate how the system is governed?		

7. Vendor & Third-Party Due Diligence Prompts

Most banks do not build their own foundation models — they license them. That dependency introduces risk that sits partly outside the bank's direct control, which makes vendor due diligence a core part of GenAI governance rather than a procurement afterthought.

Questions to Ask Foundation Model and API Providers

- What data was the model trained on, and can you disclose known limitations or gaps?
- Are our prompts and outputs logged, retained, or used to further train the model?
- How and when will we be notified of a model version change or deprecation?
- What security certifications and data-handling controls does your platform maintain?
- What is your incident response process if the model produces harmful or incorrect output at scale?
- Can we audit or independently test the model's behavior before and after updates?

Concentration Risk and Service Continuity

Relying on a single vendor for a critical GenAI workflow creates concentration risk similar to any other critical third-party dependency. Consider:

- What is the fallback plan if the vendor has an outage or discontinues the model?
- Are multiple GenAI use cases dependent on the same underlying provider, compounding exposure?
- Does the vendor contract include service-level commitments appropriate to the use case's risk tier?

Example — Vendor risk: *A bank runs three separate GenAI applications — document summarization, an internal chatbot, and a code-review assistant — all on the same foundation model provider. When that provider experiences a multi-hour outage, all three go down simultaneously, revealing a concentration risk that wasn't visible when each use case was reviewed in isolation.*

8. Building Internal Capability

Frameworks and checklists only work if an organization has the structure and skills to apply them consistently. This section outlines how to build that capability.

Suggested Governance Structure

- GenAI governance committee — cross-functional group (risk, compliance, technology, legal, business owners) that reviews and approves new use cases against the risk tiers in Step 2.
- Model risk / validation team — extends existing MRM expertise to cover GenAI-specific testing (Step 4).
- Use case owners — accountable for maintaining the intended-use documentation (Step 1) and escalating changes in scope.
- Data governance liaison — ensures Step 3's data questions are answered before any new integration goes live.

Roles and Training

Existing model risk and compliance staff typically need targeted upskilling rather than a wholesale new function — most MRM principles transfer directly, but staff need fluency in GenAI-specific failure modes like hallucination and prompt sensitivity.

- Structured training on GenAI concepts, applications, and risk considerations helps build this fluency without requiring a data science background.
- Certification programs — such as GSDC's Certified Gen AI in Finance and Banking Certification — can give risk, compliance, and product staff a shared vocabulary and structured knowledge base for evaluating GenAI opportunities and risks together.

The goal is not to turn every risk officer into a machine learning engineer. It's to ensure the people approving and monitoring GenAI use cases understand enough about how the technology behaves to ask the right questions.

GSDC
Global Skill Development Council

CERTIFIED AI CONSULTANT

THE AI CONSULTANT CAREER ROADMAP GUIDES YOU FROM AI FUNDAMENTALS TO PRACTICAL PROJECTS, CERTIFICATION, AND CONSULTING ROLES.

ABOUT GSDC CERTIFICATION

EBOOK

Extensive and exclusive Ebook created by world's experts to help you with understanding core concepts.

CREATED BY EXPERTS

GSDC certifications are created and authored by world's leading experts in the field.

EBOOK

Extensive and exclusive Ebook created by world's experts to help you with understanding core concepts.

LEARNING MATERIALS

Get access to learning materials such as videos, ebooks, templates, and practice exams, which will help you clear the certification exam.

LEARNING OBJECTIVE

- Understand AI capabilities, limitations, risks, and business applications while identifying the right use cases.
- Develop the skills to take AI initiatives from discovery and pilot to successful implementation and scale.

Enroll now with the code **LEARN20** To avail **20%** discount

Enroll Now

www.gsdcouncil.org