KPMG

Building an AI framework strategy for cognitive business assurance

May 2026

kpmg.com/in

KPMG. Make the Difference.

Table of contents

1	Foreword	03
2	Chapter 1: Aligning AI goals to BA objectives	05
3	Chapter 2: Designing an overarching data management strategy	08
4	Chapter 3: Modernise the datalake capabilities	16
5	Chapter 4: Conceptualise the journey of the cognitive control framework	18
6	Chapter 5: Leverage existing assurance technology stack	24
7	Chapter 6: Deploy agents with agility	26

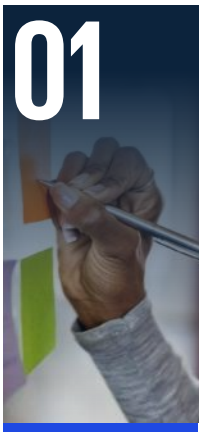
Foreword

In August 2025, KPMG in India published its Revenue Assurance and Fraud Management (RAFM) survey report, offering a market-led view of how artificial intelligence is currently being adopted across RAFM functions. While the survey revealed that organisations are directing, on average, 36 per cent of RAFM budgets toward AI, machine learning, and intelligent automation initiatives, only 14 per cent of participating mobile network operators (MNOs) are realising value at scale from these investments. More notably, 70 per cent of respondents have yet to establish data-lake readiness for AI and ML deployment, highlighting a persistent gap between ambition and execution and underscoring the absence of a structured transformation roadmap within the RAFM domain.

This thought leadership introduces cognitive business assurance: a pragmatic, future-oriented framework designed to help RAFM leaders bridge this gap. Cognitive business assurance enables organisations to systematically design, build, deploy, and govern AI-driven capabilities across the revenue assurance and fraud management ecosystem. Through a 6-step, top-down approach, this thought leadership outlines how MNOs can initiate a cognitive business assurance program that aligns enterprise strategy with compliance and control objectives. The framework supports informed decision-making across technology selection and high-value use cases, helping telcos move confidently from experimentation to scalable, sustainable impact.

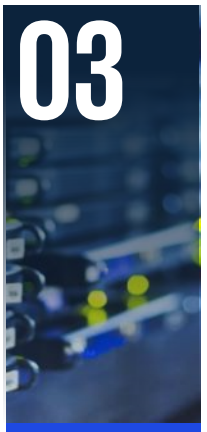
Foreword

Six chapters to building an AI framework strategy for cognitive business assurance

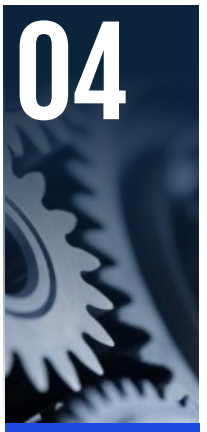


Align AI goals to BA objectives

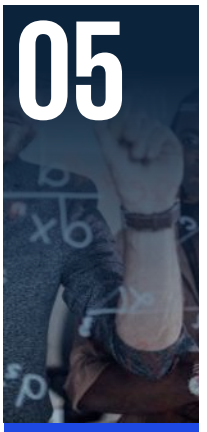
Design an overarching data management strategy



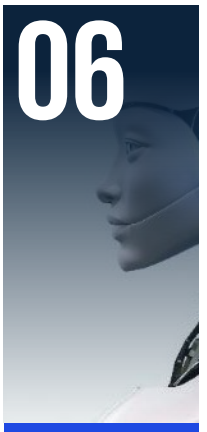
Modernise the data lakes capabilities



Conceptualise the journey of the cognitive control framework



Leverage existing assurance technology stack



Deploy AI agents with agility

As mobile network operators accelerate their adoption of artificial intelligence, the focus is shifting from experimentation to value realisation. Leadership teams increasingly recognise that deploying AI at scale particularly across Revenue Assurance and Fraud Management demands more than advanced tools and incremental automation. It requires a deliberate, enterprise-wide approach that aligns strategic ambition with data readiness, governance discipline, and execution agility. This thought leadership sets out a structured six-step framework to help organisations embed Cognitive Business Assurance as a core capability. By aligning AI initiatives with business objectives, modernising data and technology foundations, and operationalising agent-driven controls with agility, telcos can move beyond isolated use cases to build resilient, intelligence-led assurance ecosystems. The perspectives shared are designed to support confident decision-making enabling leaders to translate AI investments into sustainable outcomes, stronger controls, and measurable business impact.

Lawrence Amadi
Partner and Head of TMT
KPMG in Africa

Rahul Hakeem
Partner GRCS, TMT
KPMG in India

Arjun Malhotra
Partner GRCS, TMT
KPMG in India

Willis Genga
Partner Technology Assurance
KPMG in Kenya

Chapter 1

Aligning AI goals to BA objectives

Chapter 1:
Aligning AI goals to BA objectives

What is the global sentiment of telcos towards AI's role in transforming the BA Programme?

Of the telcos globally have apportioned budgets for AI, ML and RPA driven business assurance

Of the telcos globally believe that AI will drive transformation in RAFM operations despite current hurdles

Telcos have an increased awareness for AI and its potential in transforming BA programs. However apportioning budgets for tech investments remains a challenge

Source: KPMG's Revenue Assurance and Fraud Management (RAFM) Survey Report, 2025

While belief in the transformative or terraforming potential of AI is strong across the telecom sector, adoption remains constrained by structural and organisational realities.

Recent insights from KPMG's Revenue Assurance and Fraud Management (RAFM) Survey report published in August 2025 indicate that 76 per cent of global telcos view AI as a critical enabler for the next generation of Business Assurance programs, yet only 30 per cent have secured the budgetary commitment required to drive meaningful transformation.

This often has led to competitive business assurance programs in telcos to become laggard, resulting in sub-optimal performance of the function in mitigating revenue risks and exposures

The disconnect is largely driven by three persistent challenges:

- a short-term managerial lens that limits the AI strategy to encompass Business Assurance requirements
- capability gaps within BA teams that hinder the development of robust, value-led AI business cases for existing control frameworks
- an underlying scepticism around AI's ability to deliver consistent, accurate outcomes in time-sensitive operational environments. Addressing these barriers is essential if telcos are to convert conviction into action and unlock the full potential of AI-led business assurance.

Chapter 1:
Aligning AI goals to BA objectives

Adopting a 'Build Light' versus 'Invest Heavy' strategy...

Mobile network operators have begun to understand that "the one size fits all" mantra is no longer the technology stack landscape for the business assurance programme.

Build light when...

1. Management's investment decisions heavily rely on proof of concepts (POCs) conducted by BA teams
2. There is a strong tech team (e.g., python developers with AI experience) that can closely work with business to develop a cognitive layer over the existing stack
3. Business can leverage large language models (LLM) over an enterprise grade cloud environment

Invest heavy when...

1. Management has a pool of budget allocated for the BA teams on exploratory AI projects
2. The BA team has a trusted technology partner with containerized services to easily set up an AI layer over the existing stack
3. The organisation has a strong bend towards data privacy and has a data lake with big data capabilities for an overarching "on-the premises" AI program to run

A common misconception within business assurance programs is that introducing AI into the assurance layer necessitates a total replacement of the existing technology stack. In practice, this need not be the case.

For small to mid-sized mobile network operators that prefer a measured and risk-aware adoption path, a "build-light" approach can provide a pragmatic starting point. Widely available tools such as IPYBN styled notebooks for Python-based development or secured, enterprise-hosted notebook environments can serve as effective platforms for exploratory AI initiatives within the Business Assurance function.

These environments can be readily layered on top of existing data warehouses, enabling in-house technology teams to establish sandboxed test environments without disrupting core systems. As confidence and capability mature, organisations can progressively scale successful proofs of concept from lighter-weight implementations into more industrialised, "invest-heavy" deployments, extending AI applications to larger and more complex datasets. This phased transition allows business assurance teams to empirically validate which use cases deliver the greatest value within their operating context, while giving leadership the evidence required to justify broader investment. In doing so, telcos can balance innovation with control building momentum toward cognitive assurance without over-committing upfront.

Chapter 2

Designing an overarching data management strategy

Chapter 2:
Designing an overarching data management strategy

Are ROI on tech investments and data governance correlated

2x

with a combination of readiness, governance, agility, and execution discipline, it was observed that that the average technology Return On Investments (ROI) stands at 200 per cent

45%

of 2500 high performance driven global tech organisations expressed that protecting the data of the organisation using controls aligned to security standards, effective governance processes, and enabling security technologies was critical to their AI strategic goals

High performing telcos and global tech organisations in KPMG's global tech report have conveyed that streamlined governance has resulted in higher returns on tech investments

Source: KPMG Global Tech Report 2026

Agentic AI systems are inherently data-centric, where the quality, lineage, and governance of data directly influence system behaviour and decision-making. Recent research emphasises a transition from model-centric to data-centric AI, highlighting that governance must evolve accordingly to support the needs of autonomous pipelines.
A robust data governance framework forms the foundation of any trustworthy agentic AI system. Unlike traditional governance models that enforce rules after data has already been used, agentic AI workflows require controls that are active at every stage from the moment data enters the organisation to the point where an agent uses it to make a decision. This means governance must operate at ingestion (checking that data comes from trusted sources and is properly labelled), at processing (ensuring that sensitive data is not used without the right protections, such as anonymisation, redaction, encryption etc.), and at consumption (making sure that agents only access the data they are permitted to use, in real time).

Chapter 2:
Designing an overarching data management strategy

A layered, three-tier governance architecture has emerged as the dominant approach for achieving this. At the ingestion layer, governance ensures that data originates from trusted sources and is accompanied by rich metadata. The processing layer enforces rules during transformation and model interaction, such as preventing the use of sensitive data without anonymisation. The consumption layer governs how agents access and utilise data in real time, embedding compliance checks directly into execution workflows. This layered approach enables scalable enforcement of policies while maintaining the flexibility that autonomous agents require.

Governance is increasingly implemented as 'policy-as-code' a practice where governance rules are written as software rather than documented as procedures. This eliminates the need for manual audits by allowing rules to be automatically applied every time data moves or is used. Complementary components such as metadata stores, policy engines such as an open policy agent), and audit loggers provide the transparency, accountability, and regulatory alignment needed for trustworthy AI operations.

A globally recognised structure for organising all of these data management activities is provided by the Data Management Association (DAMA) Data Management Body of Knowledge (DMBOK), which defines 11 knowledge areas that together cover every aspect of how an organisation manages its data. While DAMA was originally designed for traditional enterprise data management, each of its 11 areas has direct relevance and requires specific adaptation for AI-driven environments. In simple terms: DAMA provides the blueprint; AI requires the blueprint to be updated and automated.

Data governance the practice of setting rules for how data is owned, used, and protected must now extend to cover AI models themselves: who is accountable for a model's decisions, how those decisions are logged, and how policy violations by an autonomous agent are detected and corrected. Data quality management, which traditionally focused on fixing errors before loading data into a warehouse, must become a continuous process: checking that data is still accurate, complete, and relevant at the moment an agent uses it, not just when it first arrived. Metadata Management recording information about data assets is now essential for tracing exactly which data was used to train a model, in what version, and under what conditions, so that decisions can be explained and audited.

Data architecture must evolve from siloed data warehouses to 'lakehouse' designs that can simultaneously serve historical analytics and live agent queries, supported by feature stores that cache pre-engineered signals for fast, consistent model serving. Data security controls must address AI-specific risks: ensuring that personal information is removed from training data, that access to sensitive datasets is role-governed, and that differential privacy techniques are applied where individuals could be re-identified. Data integration practices must fuse feeds from multiple source systems in real time for example, combining CDR feeds from network nodes using event streaming platforms so that agents always work with a consistent, up-to-date view of the data landscape.

Are data-lakes being leveraged enough by AI-led BA programs in telcos?

of the leading telcos have ensured data lake readiness for their AI and ML deployments

of the other telcos state that while they do have a data-lake, it is not utilised for AI and ML Deployments

Verdict: Majority of telcos are yet to scale the power of their enterprise data lakes

Source: KPMG's Revenue Assurance and Fraud Management (RAFM) Survey Report, 2025

Chapter 2:
Designing an overarching data management strategy

The remaining DAMA domains - Data warehousing and BI, reference and master data, document and content management, data modelling and design, and data storage and operations each play supporting roles: from maintaining canonical subscriber identifiers that enrich model training data, to managing the storage lifecycle of AI artefacts such as prompt templates and model configuration files, to ensuring that cloud-native storage tiers serve recent data at low latency while archiving older records cost-efficiently. The table below maps all 11 DAMA knowledge areas to their specific AI-era enhancements:

Applying the principles of DAMA to practices in for AI in data lake environments

DAMA Knowledge area	What this means in practice for AI in data lake environments?
Data governance	The rules and policies that control how data is used are now written as code and enforced automatically so every AI action is logged and compliant without manual checks.
Data quality	Rather than cleaning data just once before loading it, quality is now checked continuously throughout a pipeline catching drift, errors, and stale values before they reach a model.
Metadata management	Every dataset used to train a model carries a 'birth certificate': where it came from, how it was transformed, what version it is, and what the model is allowed to do with it.
Data architecture	Lakes and warehouses are merged into 'lakehouse' designs that serve both historical analytics and live agent queries; feature stores cache pre-engineered signals for fast model serving.
Data security	AI-specific controls are applied to training data personal information is anonymised, access to sensitive datasets is governed by role, and differential privacy techniques protect individuals.
Data integration	CDR feeds from multiple network nodes are fused and harmonised in real time using event streaming (e.g. Kafka) so that agents always see a consistent, up-to-date view of subscriber activity.
Data warehousing and BI	Reporting is increasingly automated: agents surface anomalies directly into dashboards and generate narratives, replacing manual analyst review cycles with near-real-time insight delivery.
Reference and master Data	Canonical subscriber identifiers (IMSI, MSISDN) and network topology records are maintained as master data, enriching CDRs with consistent dimensional context for model training.
Document and content management	AI artefacts model configuration files, prompt templates, audit logs, and agent decision records are stored, versioned, and retrievable as first-class managed documents.
Data modelling and design	Flexible schema-on-read designs accommodate the variability of CDR data across network generations, while vector stores enable embedding-based retrieval for LLM-powered agents.
Data storage and ops	Cloud-native object stores with tiered hot/warm/cold storage ensure that recent CDRs are available in milliseconds for real-time agents, while older records are archived cost-efficiently.

Chapter 2:
Designing an overarching data management strategy

4 Pillars for assessment of data-lake readiness for AI

4. Meta data richness
Assesses whether datasets carry sufficient metadata for model cards, bias audits, and regulatory reporting

1. Data quality fitness for machine learning
Assess if datasets meet the completeness, accuracy, and freshness thresholds that model training requires

3. Governance automation
Assesses the degree to which policies are enforced programmatically rather than through manual review

2. Lineage and provenance
Assess if every dataset used in training or inference can be traced back to its origin without gaps

Organisations that do not receive a strong core in even one of these dimensions have an elevated risk of unreliable AI and compliance failures. While DAMA-DMBOK established the foundational vocabulary of data governance, three subsequent frameworks have advanced the discipline specifically in the context of AI, cloud-native architectures, and regulatory accountability. The table below summarises each framework's key advancements beyond DAMA:

How data ready are tech organisations globally for their AI Strategy?

36% of the leading telco and tech companies state that quality of organisational data is a challenge for AI strategy

Verdict: A majority of telcos and tech-giants are yet to realise the impact of data management on their enterprise AI strategy

Source: KPMG Global Tech Report 2026

Chapter 2:
Designing an overarching data management strategy

of the leading telcos globally are yet to embark on a data management strategy for their business assurance programmes

'Majority of telcos globally still lack well defined approaches to adopting methodologies for Data Pre-Processing, Cleaning, and standardization'

High-quality data is a prerequisite for reliable agentic AI, yet static pre-processing pipelines are insufficient in dynamic environments. Research consistently identifies accuracy, completeness, consistency, timeliness, and validity as the five critical dimensions of data quality that determine whether a dataset is fit for model training. These dimensions must be monitored continuously not just at ingestion because data quality degrades over time due to network changes, upstream system updates, or integration failures.

Modern approaches advocate for a continuous data quality lifecycle, where data is evaluated not only when it first enters the pipeline but also at the point of use. This is particularly important for agentic systems that consume data from diverse and rapidly evolving sources. Embedding quality checks within pipelines ensures that agents always operate on reliable, up-to-date data.

Comprehensive pre-processing involves handling missing values, detecting and removing outliers, normalising numerical features, and encoding categorical variables all of which are foundational steps for machine learning readiness. For unstructured data, text cleaning and transformation techniques are essential for downstream processing by language models.

The complexity of modern data environments necessitates adaptive pre-processing systems. Automated machine learning approaches enable pipelines to dynamically select and apply the most appropriate pre-processing technique based on the current characteristics of the data. These systems reduce manual intervention and improve scalability, making them well-suited for agent-driven architectures where data characteristics can change without warning.

Best practices emphasise reproducibility and explainability: Every transformation should be logged with its data version, logic, and timestamp, enabling organisations to audit and validate preparation decisions. This is particularly critical in regulated contexts such as telecommunications where pre-processing choices can materially influence model outcomes and where regulators may require evidence that training data was handled appropriately.

Chapter 2:
Designing an overarching data management strategy

Data management for AI based business assurance controls

For telecommunications environments, Call Detail Records (CDRs) form the primary dataset. These records span multiple source systems MSC/MSS, SBC, SGSN/GGSN, PGW, OCS, Mediation, Billing, and 5G UPF each with distinct schemas, processing modes (real-time vs. batch), and quality characteristics. The table below provides recommended machine learning approaches, data preparation methodologies, and model tuning strategies mapped to each CDR source and its associated use cases:

Use case	CDR source(s)	Recommended algorithm / Boosting Method	Data preparation and training methodology	Model tuning approach
Fraud detection (SIM Box, VoIP, Interconnect)	MSC/MSS, SBC, IGW/GMSC	Gradient boosting (XGBoost, LightGBM) captures non-linear call-pattern anomalies; Isolation Forest for unsupervised SIM cluster outliers	SMOTE oversampling on minority fraud class; feature engineering on call duration ratios, trunk group deviation, ASR drops; deduplicate CDRs before training	Rolling-window feature store (15-min / 1-hr buckets) per IMSI / MSISDN; near-real-time scoring pipeline; retune class_weight after each fraud campaign
Roaming fraud and data leakage	SGSN, GGSN, Roaming Platform (TAP/NRTRDE)	Random Forest + SHAP explainability; LSTM for session-level temporal patterns across roaming legs	Normalise data volume and session duration; encode APN as categorical; impute missing roaming records with per-operator-pair median	Quarterly fine-tune on recent TAP/NRTRDE files to capture new roaming-partner behavior drift; PSI monitoring on APN feature distribution
Revenue assurance and rating leakage	PGW, Billing System, mediation	XGBoost Regressor for expected-vs-actual revenue delta; rule-augmented anomaly scoring on zero-charge and missing-rate flags	Standardise rated-amount distributions; engineer discount-leakage ratio feature; remove duplicates and orphan CDRs before training	Monthly retraining on rated CDR extracts; A/B holdout of last billing cycle as validation; trigger retraining when KS-test detects revenue distribution shift
Online charging and credit risk	OCS / IN	LightGBM binary classifier for credit risk; Prophet / LSTM time-series forecasting for unit-grant depletion rate	Balance-before/after delta as primary feature; encode granted-unit types; mini-batch training to handle real-time streaming CDRs	Incremental online learning to adapt to new prepaid plan structures without full retraining; monitor unit-grant error rate as a proxy drift signal
5G Slice charging leakage	UPF (5G)	CatBoost multi-class classification across slice types; contrastive learning for SUPI-level slice usage profiles	One-hot encode DNN / QoS Flow; normalise Slice RA fields; augment sparse 5G records with CTGAN synthetic slice-usage data	Transfer learning from 4G PGW model fine-tuned on UPF data to address 5G data scarcity; retrain as 5G subscriber volumes grow
Interconnect and Wholesale settlement	Interconnect Billing System, Wholesale MVNO mediation	DBSCAN clustering to surface partner-level settlement anomalies; rule-augmented ML for contractual threshold breaches	Aggregate CDRs to partner-day grain; engineer settlement-deviation percentage as label; forward-fill missing duration within partner sessions; enrich with contract metadata	Batch monthly retraining aligned to billing cycle; use partner contract metadata as contextual features; flag anomalies exceeding 2σ of historical settlement variance
Payment gateway transactions and anomaly identification	Data from payment gateways and third-party transactions	Machine learning based pattern identification and link analysis to identify potential fraudulent payment	Feed potential fraudulent transactions in the past to machine learning algorithms and link suspicious user activities to suspicious payments	Define batch procedures and window periods for the ML algorithm to run through implementation of fuzzy lookups at periodic intervals

Chapter 2:
Designing an overarching data management strategy

Is feature engineering the new AI building material for BA?

Feature engineering should operate at IMSI, MSISDN, and partner-grain rather than at raw event level, using time-bucketed aggregations (15-minute, 1-hour, and daily windows) to capture behavioural patterns. Class imbalance endemic to fraud and leakage datasets should be addressed through oversampling techniques such SMOTE, ADASYN, or cost-sensitive learning applied consistently. Concept drift, driven by new tariff plans, roaming agreements, or 5G slice configurations, should be monitored using Population Stability Index and KS-tests on feature distributions, with automated retraining pipelines triggered when drift thresholds are brhed. Finally, every CDR transformation must be logged and explainability tools such as black-box machine learning models like SHAP and LIME should be embedded in agent decision outputs to support dispute resolution and regulatory audit.

Transition into distributed frameworks...

The transition to Agentic AI requires a rethinking of system architectures, moving from monolithic designs to modular, distributed frameworks. A reference architecture for enterprise agentic systems outlines five core building blocks:

01 Orchestrator

Plays a central role, decomposing high-level objectives into executable tasks and coordinating the activities of specialised agents.

02 Specialised Agents

Designed to perform specific functions such as data pre-processing, anomaly detection, risk assessment, compliance checking, and reporting

03 Tool interfaces

Allow agents to interact with external systems including databases, APIs, and analytical engines, extending functionality beyond static reasoning.

04 Memory Systems

Encompassing both short-term context and long-term knowledge bases, supporting continuity and learning across agent interactions.

05 Comm. Infrastructure

Facilitates coordination between agents, enabling asynchronous and distributed execution essential for complex, multi-step workflows.

Chapter 3

Modernise the data lake capabilities

Chapter 3:
Modernise the data-lake capabilities

Traditional data platforms, designed primarily for batch processing, are ill-equipped to handle the real-time demands of Agentic AI systems. Modern architectures must support low-latency data access, high-throughput processing, and dynamic scalability to serve autonomous agent workloads effectively.

Regular assessment of AI workloads is key...

Ongoing assessment of capacity and infrastructure is essential to ensuring that the data lake environment remains fit for purpose as agentic workloads evolve. Recent research proposes predictive autoscaling frameworks that leverage time-series forecasting and telemetry data to anticipate workload demands. By provisioning resources proactively, these systems ensure consistent performance while optimising cost efficiency, particularly important for agentic workloads that may exhibit bursty and unpredictable patterns.

The integration of telemetry, forecasting, and automation creates a resilient infrastructure capable of supporting continuous agent operations. By aligning infrastructure capabilities with the needs of autonomous systems, organisations can ensure reliability, scalability, and responsiveness.

Compute and Storage Capacity

Review current utilisation against projected demand, identifying bottlenecks or over-provisioned resources.

Latency and throughput

Evaluate end-to-end pipeline performance to ensure data is available to agents within required time constraints.

Infrastructure health

Assess the availability of anomaly detection, automated diagnostics, and remediation mechanisms.

Scalability readiness

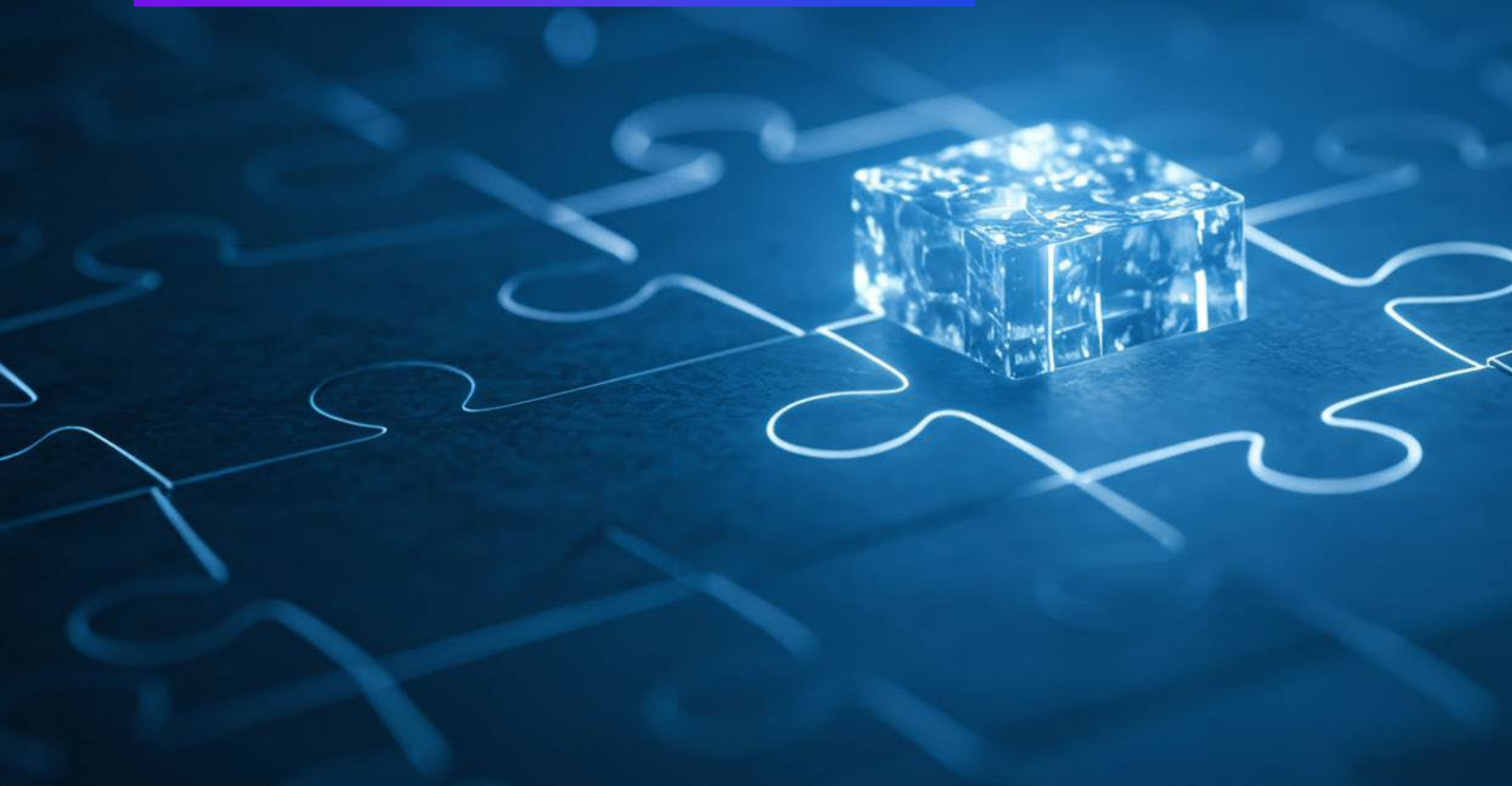
Ensure that dynamic scaling capabilities are appropriately configured and tested for anticipated load scenarios.

How to combat IT infra challenges for AI?

1. Use telemetry and forecasting outputs to communicate capacity needs, ahead of demand spikes
2. Implement automated alerting for IT performance degradation or other infrastructure anomalies
3. Establish clear processes for escalating critical infrastructure gaps such as storage exhaustion or network bottlenecks
4. Align data lake modernisation roadmaps with IT planning cycles and factored into IT Investment plans
5. Maintain clear, version-controlled documentation of infrastructure requirements

Chapter 4

Conceptualise the journey of the cognitive control framework



Chapter 4:
Conceptualise the journey of the cognitive control framework

Conceptualising a "one-AI agent" strategy

The idea of using one Agentic AI across multiple teams originates from the nature of Revenue Assurance and fraud management itself. Revenue leakage does not belong to a single department. It originates in customer behavior, passes through products and services, traverses network resources, is processed by mediation and billing systems, and finally manifests as a financial outcome. Despite this being one continuous economic flow, organisations operate it through separate teams – business assurance, revenue assurance, billing, IT, network, finance, audit and compliance.

Each team traditionally views the same issue through a local lens. The one agent model is designed to preserve a single logical understanding of the revenue lifecycle while allowing the output to be contextually presented to different teams. This is the core reason one agent is preferred over multiple agents. Following are the key characteristics of a "one-agent" strategy

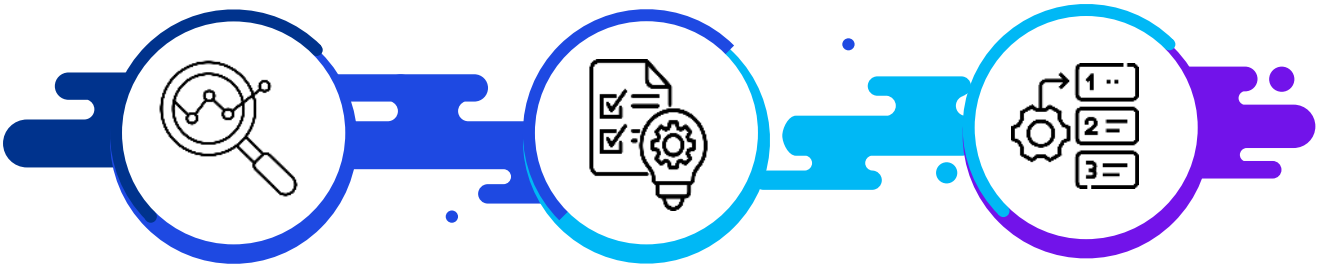
Single reasoning model

A shared understanding of revenue entities

Common causal logic

Unified historical memory

Internal structure: Three layers of a "one-AI BA agent"



Data observation layer

Establishes a neutral, trusted foundation by ingesting raw operational data end-to-end, preserving integrity and context without applying business interpretation.

Causal reasoning layer

Reconstructs the revenue lifecycle by explaining why deviations occur—linking intent, execution, and outcomes to deliver trusted root-cause insights.

Economic prioritisation layer

Ranks identified issues by financial impact, persistence, customer exposure, and recoverability to focus action on material business risk

Chapter 4:
Conceptualise the journey of the cognitive control framework

Illustrative examples of a "one-AI agent" being used across multiple O/BSS teams:

Revenue Assurance and fraud management teams

For the Revenue Assurance team, the agent focuses on identifying systemic issues rather than isolated incidents. It highlights recurring leakage patterns, classifies root causes, and evaluates the effectiveness of existing . The agent:

- Identifies recurring types of revenue leakage
- Maps these issues to missing, weak, or ineffective controls
- Recommends preventive actions to avoid recurrence.

In this context, the RA team is not exposed to raw system data or logs. Instead, they see consolidated patterns, quantified risks, and control-oriented insights that support proactive assurance.

Billing teams

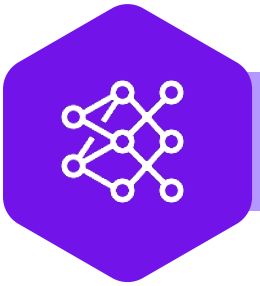
When serving the Billing team, the agent shifts its emphasis toward pricing execution and billing accuracy. It explains how configured tariffs and discount rules were applied and where pricing logic deviated from intended behavior. The agent:

- Highlights incorrectly applied rates or discounts
- Explains failures in rating or tariff logic
- Simulates corrected billing outcomes based on intended pricing rules

Billing teams receive configuration level explanations derived from the same reasoning chain used elsewhere, allowing them to correct issues without re investigating the problem from scratch.

Chapter 4:
Conceptualise the journey of the cognitive control framework

Illustrative examples of a "one-AI agent" being used across multiple O/BSS teams:



IT and Network team

In the IT and Network context, the agent focuses on data integrity and technical behavior across systems. It correlates network, mediation, and configuration data to explain how technical issues translate into business impact. The agent:

- Identifies missing, delayed, or malformed usage records
- Connects technical failures to downstream financial effects
- Avoids attributing faults without clear, traceable evidence.

IT teams, therefore, see service and resource level correlations rather than revenue metrics, enabling targeted technical remediation grounded in factual impact.

Finance team

For Finance, the agent presents a financially grounded view of the same issues, emphasising clarity and certainty rather than anomaly detection. The agent:

- Calculates recoverable and non recoverable revenue
- Differentiates between adjustments and write offs
- Tracks recovery performance over time.

In this context, Finance teams see clear monetary impact and recovery status, allowing confident accounting treatment and management reporting without navigating technical complexity.

Chapter 4:
Conceptualise the journey of the cognitive control framework

Applying the "one-agent" strategy to different data sources

S.No	CDR / Source Node	Key Fields	Use Cases	Processing Mode	Field Criticality	Agent Narrative (How One Agent Uses This Source)	Typical Agent Decisions
1	MSC / MSS	IMSI, MSISDN, Call Start Time, Call Duration, Trunk Group	SIM Box Fraud, Interconnect Fraud, Rating Leakage	Near Real-Time / Batch	Mandatory	Agent studies switch-level calling behavior to detect abnormal termination, correlating it with charging and revenue loss.	Flag SIM clusters, recommend barring, calculate revenue loss
2	International Gateway Switch (IGW/GMSC)	Origin Operator, Route ID, Call Duration, ASR	SIM Box Fraud, Interconnect Bypass	Near Real-Time / Batch	Mandatory	Agent correlates international ingress routes with local termination to detect bypass of interconnect charging.	Recommend route blocking, support dispute cases
3	Session Border Controller (SBC)	SIP Call ID, Source IP, Call Setup Time, Duration	SIM Box Fraud, VoIP Fraud	Near Real-Time	Mandatory	Agent analyses SIP signaling to identify VoIP gateways masquerading as subscriber calls.	Identify illegal termination paths
4	SGSN	IMSI, Cell ID, Session Duration, Data Volume, APN	Roaming Fraud, Data Leakage	Batch	Mandatory	Agent reviews packet session behavior to detect roaming misuse or abnormal consumption.	Trigger roaming controls or alerts
5	GGSN	IMSI, APN, Assigned IP, Volume, Session Time	Data Revenue Assurance, Promo Abuse	Batch	Mandatory	Agent compares data usage at gateway level with charged volumes to ensure correct monetisation.	Recommend tariff or promo corrections
6	PGW	IMSI, Rating Group, Volume UL/DL, Charging ID	Billing Leakage, Bill Shock	Near Real-Time / Batch	Mandatory	Agent monitors real-time data charging behavior to prevent undercharging or bill shock.	Trigger alerts or charging corrections

Chapter 4:
Conceptualise the journey of the cognitive control framework

Applying the "one-agent" strategy to different data sources

S.No	CDR / Source Node	Key Fields	Use Cases	Processing Mode	Field Criticality	Agent Narrative (How One Agent Uses This Source)	Typical Agent Decisions
7	UPF (5G)	SUPI, Slice ID, DNN, Volume, QoS Flow	5G Charging Leakage, Slice RA	Near Real-Time	Optional	Agent evaluates slice-wise usage to ensure differentiated 5G services are monetised correctly.	Highlight slice leakage risks
8	Mediation System	Record ID, Reject Reason, Missing Attributes	Usage Mediation Failure	Batch	Mandatory	Agent detects data loss between network and billing, explaining missing or rejected usage.	Recommend mediation fixes and recovery
9	OCS / IN	Session ID, Granted Units, Used Units, Balance Before/After	Online Charging Leakage, Credit Risk	Real-Time	Mandatory	Agent tracks balance movements and unit grants to detect abnormal consumption patterns.	Trigger credit controls or fraud cases
10	Billing System	Rated Amount, Tariff ID, Discount ID, Zero Charge Flag	Rating Errors, Revenue Assurance	Batch	Mandatory	Agent validates billing outcomes against expected product pricing behavior.	Recommend billing corrections
11	Roaming Platform (TAP/ NRTRDE)	IMSI, VPLMN, Usage Timestamp, Accrued Charges	Roaming Fraud, Credit Risk	Batch	Mandatory	Agent evaluates roaming exposure growth to balance fraud risk with customer experience.	Initiate barring or notification
12	Interconnect Billing System	Partner ID, Route ID, Duration, Settlement Amount	Interconnect Revenue Leakage	Batch	Mandatory	Agent reconciles internal traffic with partner settlements to detect commercial leakage.	Support recovery and disputes

Chapter 5

Leverage existing assurance technology stack

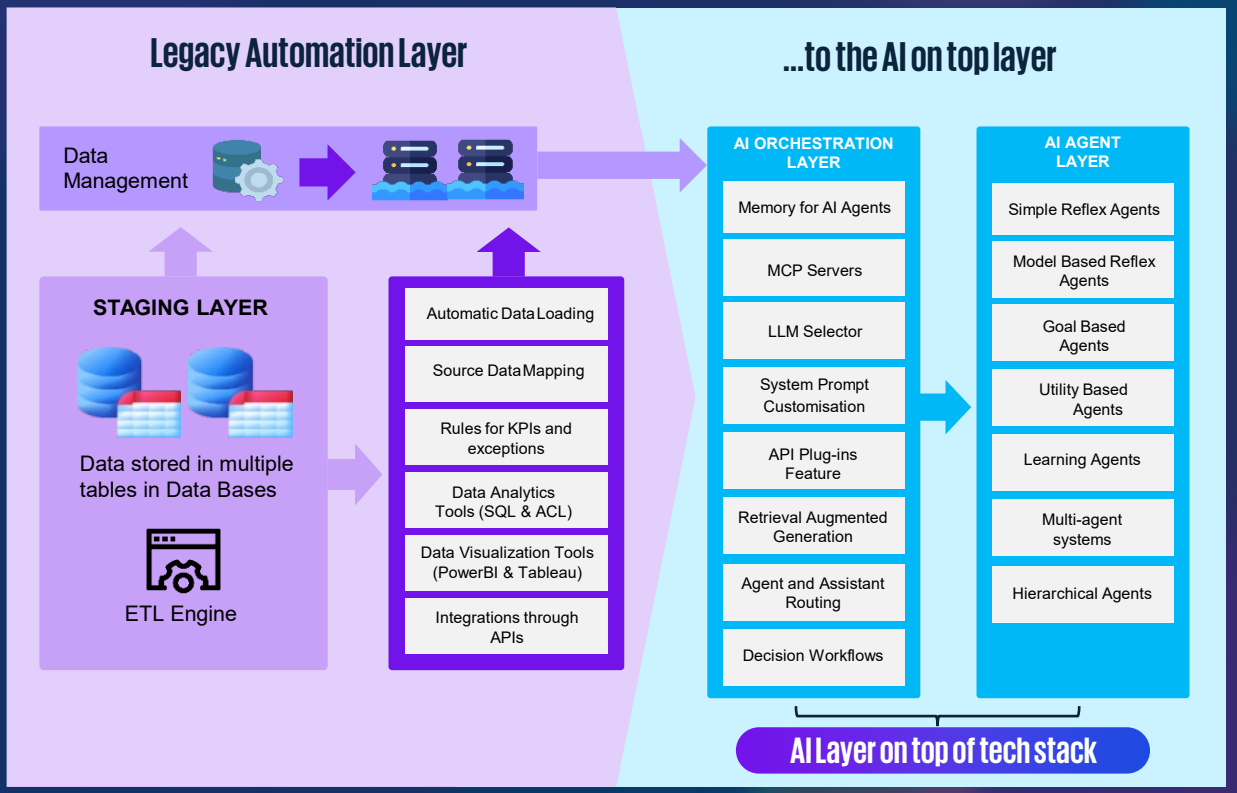
Chapter 5:
Leverage existing assurance technology stack

Adding an AI layer does not mean a total removal of the existing assurance layer

While Business Assurance teams increasingly aspire to evolve from legacy rule-based engines toward dynamic, AI-driven detection models, a common concern persists that such a shift necessitates the wholesale replacement of existing assurance technology infrastructure. This transformation need not be disruptive. Intelligent AI layers can be architected over and above the current assurance stack, leveraging existing controls, workflows, and critically rich historical assurance data. These historical datasets provide a valuable training ground for AI agents, enabling them to learn enterprise-specific patterns and contexts while operating alongside established systems. This layered approach allows organisations to modernise assurance capabilities incrementally, preserving prior investments while progressively enhancing accuracy, adaptability, and insight through AI-enabled intelligence.

The journey from automation to artificial intelligence

Organisations globally are in the process of building an AI layer over the existing automation layer for business assurance with an objective of improving the turn-around time for the audit while providing actionable management insights



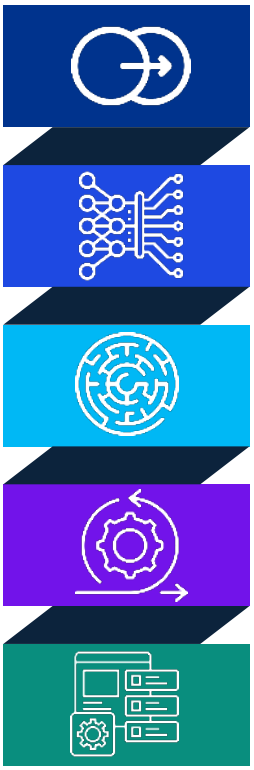
Chapter 6

Deploy AI agents with agility

Chapter 6:
Deploy AI agents with agility

Leveraging agile for BA AI Assurance agents...

Deployment of AI agents within Cognitive Business Assurance represents a fundamental shift from traditional, rule-based assurance models. Unlike static controls, agentic systems continuously evolve through interaction with enterprise datasets and operational feedback, requiring implementation strategies that support iterative refinement rather than one-time configuration. As a result, agile delivery frameworks underpinned by user-story-driven planning, controlled release management, and disciplined testing provide the governance necessary to introduce agentic monitoring safely across revenue-critical workflows while maintaining alignment between business objectives and operational data. This approach is especially critical in regulated environments, where trust, auditability, and resilience are as important as speed of innovation, enabling organisations to scale AI-driven assurance across risk, compliance, quality, and knowledge use cases without fragmentation or loss of control.



Shift from deterministic to adaptive assurance: Traditional, rule-based Business Assurance deployment models are insufficient for AI-driven, agentic assurance systems that learn, adapt, and evolve over time.

AI monitoring demands iterative deployment: Machine-learning-enabled assurance relies on continuous validation and recalibration due to dataset drift, evolving service behavior, and changing operational realities.

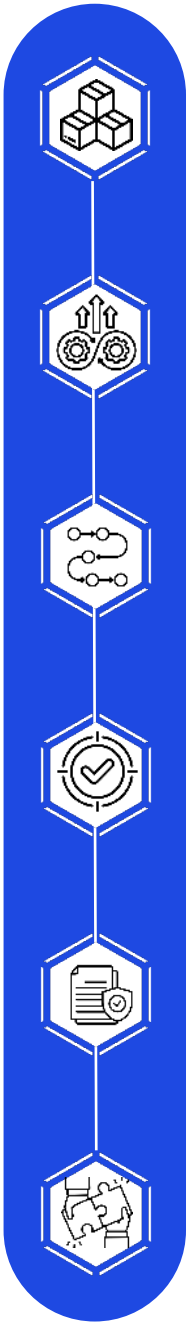
Telecom environments amplify complexity: Mediation, charging, roaming, and partner settlements generate highly heterogeneous and dynamic datasets, requiring deployment models that can adjust monitoring logic progressively.

Agile frameworks outperform waterfall models: Unlike static controls, AI agents evolve through prompt refinement, model improvement, and feedback loops—making upfront, fixed requirements impractical.

Agility enables controlled value realisation: Incremental rollout, frequent review, and feedback-driven adaptation allow organisations to introduce agentic assurance safely while ensuring accuracy, trust, and operational relevance.

Chapter 6:
Deploy AI agents with agility

Leveraging agile for BA AI Assurance agents...



Step 1: Acknowledge the limits of static deployment models
Recognise that AI assurance agents are adaptive and probabilistic by nature, requiring iterative refinement rather than one-time, rule-based configuration

Step 2: Anchor deployment on agile delivery principles
Replace waterfall rollout with sprint-based delivery to enable gradual introduction of agent capabilities, frequent review, and responsive adjustment to data and operational change.

Step 3: Roll out agents incrementally across assurance domains
Deploy agents in phased scopes—starting with selected mediation, billing, or settlement workflows—to reduce operational risk and validate performance before scaling.

Step 4: Embed continuous accuracy validation into each sprint
Continuously test and recalibrate agent behaviour against live datasets to address data drift, false positives, and evolving transaction patterns throughout deployment cycles.

Step 5: Use agile as a governance and control mechanism
Leverage short delivery cycles, evidence-based reviews, and controlled releases to ensure transparency, auditability, and resilience in revenue-critical assurance environments

Step 6: Enable cross-functional collaboration by design
Align assurance analysts and AI engineering teams through shared backlogs, sprint reviews, and iterative refinement to ensure agent behaviour reflects real business assurance intent.

Chapter 6:
Deploy AI agents with agility

Building user stories for creation of BA agents...

Step 1: Translate assurance objectives into user-centric outcomes
Clearly articulate what the organisation expects the AI agent to detect, validate, or explain by framing requirements in user stories that link business roles, desired outcomes, and measurable control objectives.

Step 2: Define agent capabilities with clarity and constraint
Use user stories to specify the data sources, operational boundaries, success criteria, and constraints under which each agent capability must operate, ensuring transparency and purpose-driven design.

Step 3: Enable progressive capability expansion through prioritised stories
Break down larger assurance goals into epics, features, and smaller stories to support incremental rollout, allowing capabilities to mature progressively as data readiness and confidence increase.

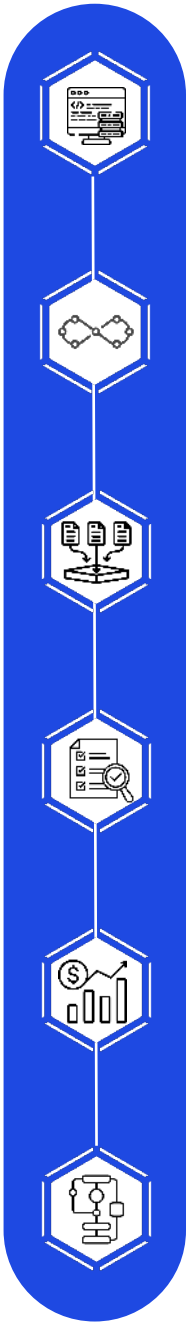
Step 4: Maintain traceability between intent and behaviour
Leverage acceptance criteria, identifiers, and performance expectations within user stories to create a clear, auditable link between assurance objectives and deployed agent behaviour.

Step 5: Sequence deployment across assurance workflows deliberately
Prioritise user stories based on business value, control criticality, and data maturity to roll out agents gradually across mediation, billing, settlement, and partner assurance domains.

Step 6: Embed governance and compliance into delivery artefacts
Capture regulatory, risk, and governance requirements directly within user stories so that transparency, auditability, and control alignment are enforced throughout the deployment lifecycle.

Chapter 6:
Deploy AI agents with agility

Deploying CI/CD pipelines and managing the SDLC lifecycle...



Step 1: Expand the SDLC mindset beyond code delivery
Recognise that AI assurance deployment must govern models, prompts, data pipelines, orchestration logic, validation artefacts, and monitoring controls—far beyond traditional application releases.

Step 2: Establish CI/CD and MLOps as core operational enablers
Implement CI/CD and MLOps pipelines to move AI agents from experimentation to production through repeatable, governed lifecycle processes aligned with evolving enterprise data environments.

Step 3: Manage co-evolution of code, models, and pipelines
Ensure deployment pipelines evolve in lockstep with changes to models, prompts, thresholds, validation logic, and dependencies to prevent drift between agent behaviour and release controls.

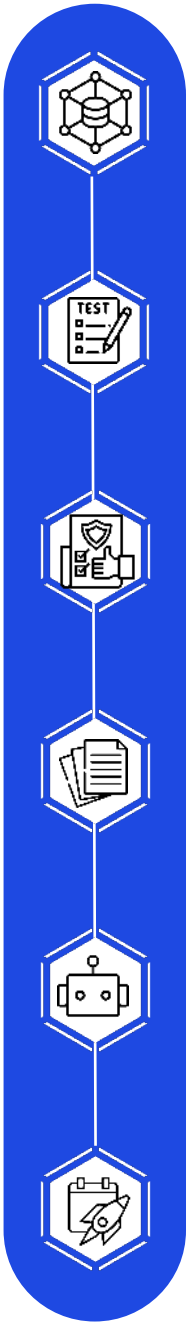
Step 4: Embed reproducibility, quality gates, and security checks
Integrate versioning, automated testing, validation thresholds, security scanning, and controlled promotion into every pipeline stage to ensure only trusted capabilities reach production.

Step 5: Use pipeline controls to protect revenue-critical operations
Apply evidence-based promotion criteria covering accuracy, reliability, safety, performance, and cost, before changes affect live assurance workflows.

Step 6: Treat CI/CD as an enterprise delivery discipline
Standardise pipelines, tooling, and release practices across teams to improve coordination, traceability, and scalability while preserving governance, auditability, and operational resilience.

Chapter 6:
Deploy AI agents with agility

Defining Test Use Cases for Model Training Quality



Step 1: Expand testing beyond standalone model accuracy
Treat testing as a lifecycle activity spanning component validation, integration testing, deployment checks, and post-deployment evaluation within live assurance environments.

Step 2: Design test scenarios around assurance-critical quality attributes
Define tests based on business-relevant qualities such as accuracy, grounding, reliability, latency, safety, scalability, and cost—rather than technical metrics alone.

Step 3: Validate training quality against enterprise-specific assurance workloads
Measure agent performance using representative Business Assurance scenarios—such as billing inconsistencies, settlement validation, and anomaly detection—rather than generic benchmarks.

Step 4: Enforce reproducible and governed evaluation practices
Standardise model versions, prompts, datasets, and scoring methods, and use sprint-level evaluation scorecards to enable repeatable, auditable, evidence-based deployment decisions.

Step 5: Test agent behaviour across multi-step assurance workflows
Evaluate how agents reason, plan, use tools, and recover from uncertainty across end-to-end monitoring journeys, not just whether they produce a correct final result.

Step 6: Gate rollout based on demonstrated operational readiness
Allow progression to wider deployment only when testing confirms robustness, consistency, and fitness for revenue-critical assurance use.

Chapter 6:
Deploy AI agents with agility

Assessment of deployment across key metrices...

Step 1: Assess technical readiness and integration fit
Evaluate architecture suitability, data quality, infrastructure maturity, and the agent's ability to integrate seamlessly with mediation, charging, settlement, and assurance platforms.

Step 2: Validate monitoring performance and drift resilience
Measure stability of anomaly detection over time, effectiveness of drift detection, and readiness of retraining, recalibration, and operational support mechanisms.

Step 3: Examine operational sustainability and lifecycle support
Assess the presence of runbooks, dashboards, alerting, evaluation routines, and ownership models required to sustain reliable performance in evolving assurance environments.

Step 4: Evaluate resource, risk, and cost viability
Confirm that staffing models, governance effort, pipeline operations, validation activities, and long-term operating costs are realistic, controlled, and economically sustainable.

Step 5: Measure organisational adoption and user acceptance
Assess usability, clarity of outputs, stakeholder trust, and the extent to which analysts and managers embed the agent into daily assurance workflows.

Step 6: Establish trust, explainability, and enterprise confidence
Validate decision traceability, interpretability of findings, transparency of limitations, and confidence in how AI outputs influence revenue protection and compliance decisions

Step 7: Gate scale-up based on holistic readiness
Advance deployment only when technical, operational, economic, and trust-based metrics collectively demonstrate fitness for revenue-critical use.

Foot notes

This thought leadership has been developed based on extensive industry research, market studies, regulatory publications, and expert insights sourced from globally recognised institutions and reports. The following references were consulted to validate the findings, trends, and recommendations presented in this paper:

- Sharma, R., et al. (2024). Data Governance for Artificial Intelligence. Springer ISIC.
- Mökander, J., et al. (2024). AI governance: a systematic literature review. AI and Ethics, Springer.
- Anderljung, M., et al. (2024). Towards Data Governance of Frontier AI Models. arXiv:2412.03824.
- Patel, N., and Singh, H. (2024). Intelligent Data Governance Frameworks: A Technical Overview. ResearchGate.
- DAMA International. (2017). DAMA-DMBOK: Data Management Body of Knowledge (2nd ed.). Technics Publications.
- EDM Council. (2021). Data Management Capability Assessment Model (DCAM) v2.0. EDM Council Publications.
- EDM Council and Cloud Providers Consortium. (2022). Cloud Data Management Capabilities (CDMC) Framework. EDM Council Publications.
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1.
- Alsheikh, A., et al. (2024). A Survey on Data Quality Dimensions and Tools for Machine Learning. arXiv:2406.19614.
- Zhang, Y., et al. (2024). Data Preprocessing Techniques for AI and ML Readiness. JMIR mHealth.
- Li, P., et al. (2024). Automated data processing and feature engineering. arXiv:2403.11395.
- García, S., et al. (2024). Data Preprocessing and Feature Engineering. MDPI Data, 6(10), 257.
- Johnson, M., and Lee, K. (2024). Agentic AI: A Reference Architecture for Enterprise Automation. IEEE ICSA.
- Wang, H., and Gupta, A. (2024). Building Blocks for Autonomous Agents: A 2024 Survey. ACM Computing Surveys.
- Chen, Y., and Patel, S. (2024). Dynamic Infrastructure Scaling for AI Agents. ACM SoCC.
- Kumar, R., and Zhao, L. (2025). Self-Aware Data Platforms: Telemetry-Driven Resource Management. CIDR.
- Software Engineering Institute (Carnegie Mellon University) Applying Agile to Machine Learning System Development https://www.sei.cmu.edu/blog/applying-agile-at-scale-for-mission-critical-software-reliant-systems/
- Machine Learning–Enabled System Development Mapping Study https://arxiv.org/abs/2506.20759
- Continuous Delivery and Automation in MLOps Pipelines https://arxiv.org/abs/2403.13115
- NIST AI Risk Management Framework https://www.nist.gov/itl/ai-risk-management-framework
- Agile Requirements Engineering for Machine Learning Systems https://arxiv.org/abs/2602.05042
- TechTarget Definition of User Story https://www.techtarget.com/searchsoftwarequality/definition/user-story
- Agile Requirements Engineering Adaptation Study https://www.sciencedirect.com/science/article/pii/S0164121225003620
- Scrum Guide (2020) https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-US.pdf
- Kreuzberger, Kühl, Hirschl Machine Learning Operations (MLOps): Overview, Definition, and Architecture IEEE Access, 2023.
- Rzig, Houerbi, Chavan, Hassan Empirical Analysis on CI/CD Pipeline Evolution in Machine Learning Projects 2024.
- Méndez, Camargo, Florez Machine Learning Operations Applied to Development and Model Provisioning Springer, 2024.
- Gupta et al. Continuous Integration, Delivery and Deployment: A Systematic Review of Approaches, Tools, Challenges and Practices Springer, 2024.
- Test and Evaluation Best Practices for Machine Learning-Enabled Systems 2023.
- Using Quality Attribute Scenarios for ML Model Test Case Generation 2024.
- Enterprise Benchmarks for Large Language Model Evaluation 2024.
- A Systematic Survey and Critical Review on Evaluating Large Language Models: Challenges, Limitations, and Recommendations EMNLP / ACL Anthology, 2024.
- AgentBoard: An Analytical Evaluation Board of Multi-turn LLM Agents NeurIPS 2024.
- Challenges with Developing and Deploying AI Models and Applications in Industrial Systems Discover Artificial Intelligence, 2024.
- ML-Enabled Systems Model Deployment and Monitoring: Status Quo and Problems SWQD 2024.
- Development and Deployment Challenges of Machine Learning Systems Springer, 2024.
- Determinants of Artificial Intelligence Adoption: Research Themes and Future Directions Information Technology and Management, 2024.
- Trust in AI: Progress, Challenges, and Future Directions Humanities and Social Sciences Communications, 2024.
- A User-Based Study on the Acceptance of Artificial Intelligence-Based Decision-Support Systems PACIS 2024.

Acknowledgements

We are extremely grateful to senior leaders from the industry, subject matter experts and KPMG team members for sharing their knowledge and insights to help us develop this report.

Authors

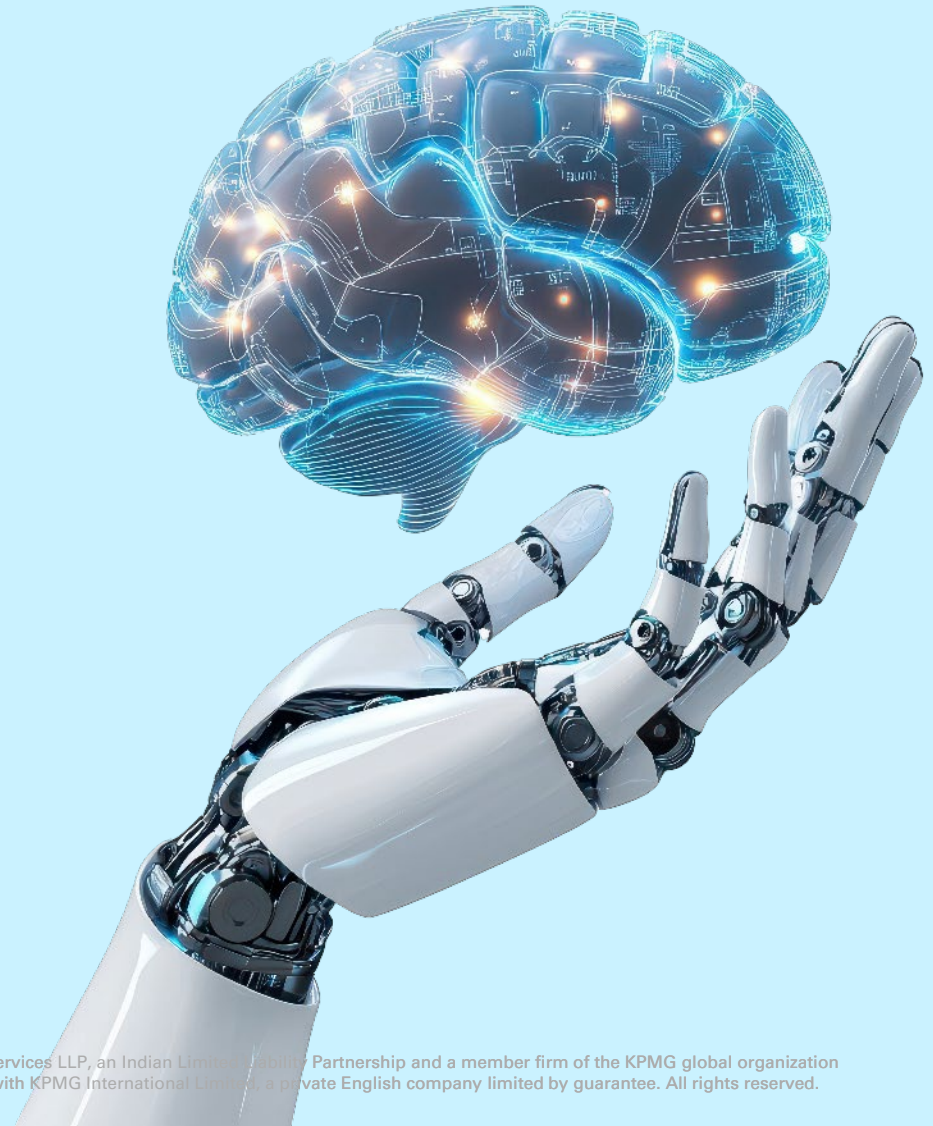
- Lawrence Amadi
- Willis Genga
- Rahul Hakeem
- Arjun Malhotra

Contributors

- Thomas Gouws
- Yatin Gaind
- Prithviraj Chauhan
- Avinash CH
- Arnab Mohapatra
- Mridula Sinha
- Catherine Mutuma
- Collins Onah
- Rohan Chaudhry
- Rahul Raj
- Deepak Kumar Hotwani
- Sanu Khetan
- Kashish Kakkar
- Khushi Mule
- Swathi Kara
- Smruthy Joseph
- Rejith G
- Devaansh Singh
- Chandresh Sharma
- Sudha Kumari
- Nishant Chechi

Design

- Arun Choudhary



KPMG contacts:

Akhilesh Tuteja
Partner and National Leader – Technology Media and Telecom
KPMG in India
E: atuteja@kpmg.com

Lawrence Amadi
Partner and Head TMT KPMG Africa
KPMG in Nigeria
E: Lawrence.Amadi@ng.kpmg.com

Thomas Gouws
Partner – Internal Audit & ERM: Jhb
KPMG in South Africa
E: thomas.gouws@kpmg.co.za

Rahul Hakeem
Partner – GRCS Telecom
KPMG in India
E: rahulhakeem@kpmg.com

Willis Genga
Partner – Audit and Technology Assurance Services
KPMG in Kenya
E: wgenga@kpmg.co.ke

Arjun Malhotra
Partner – GRCS Telecom
KPMG in India
E: arjunmalhotra1@kpmg.com

Moses Kipchirchir
Partner – Advisory
KPMG in Kenya
E: mkipchirchir@kpmg.co.ke

kpmg.com/in

Access our latest insights on KPMG Insights Edge

Follow us on:
kpmg.com/in/socialmedia

KPMG Assurance and Consulting Services LLP, Lodha Excelus, Apollo Mills Compound, NM Joshi Marg, Mahalaxmi, Mumbai – 400 011 Phone: +91 22 3989 6000, Fax: +91 22 3983 6000.