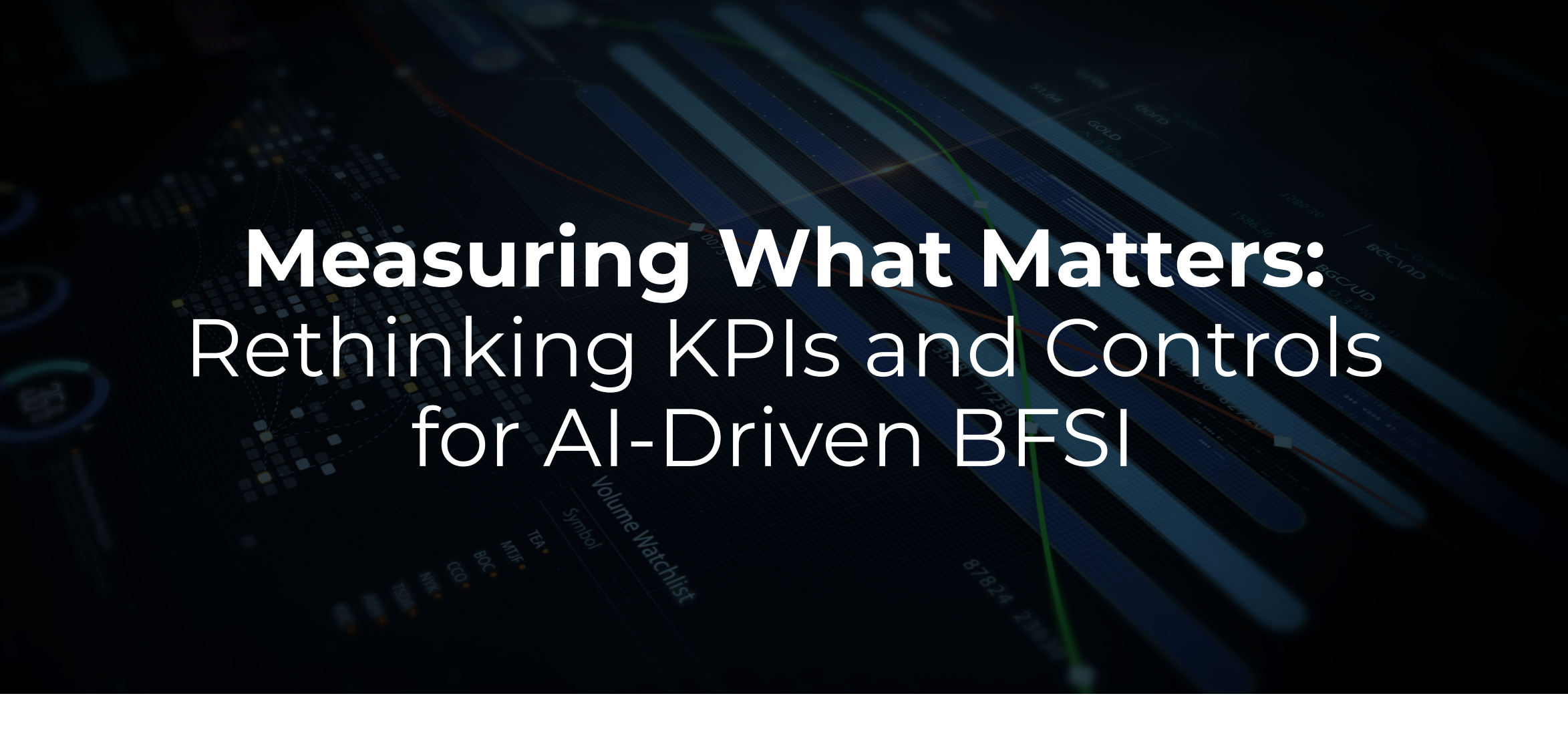

Measuring What Matters:
Rethinking KPIs and Controls for AI-Driven BFSI

How to use this paper

Use this paper as a working guide to align leadership, select priority journeys and design governance that enables safe scale.

Start with the scorecard:
Use the Executive Dashboard template for your top two or three journeys; keep it small and decision oriented.

Apply the 2×2 lens:
Ensure each journey has outcome and guardrail metrics, with leading indicators you can act on weekly (or faster for high-risk workflows).

Set decision triggers:
Pre-define thresholds that route work, increase review, tighten autonomy or pause automation.

Assign clear owners:
One accountable journey owner, plus named risk/control partners for guardrails.

Use the appendices as the build kit:
Adopt the Metrics Catalog (Appendix B) and definitions (Appendix C) to standardize measurement and evidence.

Introduction

AI is compressing decision cycles across BFSI, in fraud detection, lending, claims, onboarding, surveillance and customer service. As automation increases, small errors scale faster, and delayed intervention becomes more costly. Yet, most performance measurement frameworks remain designed for retrospective review, not real-time management.

This gap widens as organizations introduce AI-enabled and agentic workflows capable of executing multi-step decisions with limited human involvement. Leaders need metrics that answer three questions continuously: Is value being delivered? Are controls holding? And do we need to intervene - now?

Executive Dashboard: A Minimum Viable Scorecard

This executive dashboard is designed to answer one simple question: is AI improving outcomes without weakening control?

Rather than tracking dozens of indicators, this scorecard focuses leadership attention on a small set of journeys that matter most — typically two or three to start. Each journey is reviewed through a balanced lens that shows progress, risk, and confidence at the same time, enabling faster decisions about where to scale, where to intervene, and where to pause.

The intent is not performance reporting for its own sake. It is to create a regular, decision led operating rhythm that helps executives:

Confirm where AI is delivering tangible business value.

Detect early signs of control or quality drift.

Decide when to increase oversight, adjust autonomy, or safely expand automation.

The scorecard pairs value indicators (speed, cost, experience) with control indicators (stability, overrides, exceptions), ensuring that improvements in efficiency are never viewed in isolation from risk and governance.

Table 1. Scorecard Template (What to track + Decision triggers)

Scorecard block	What to track (examples)	Decision trigger
Outcome (leading)	In-flight cycle time / queue age; STP by step; digital adoption; always-on sentiment	If Leading outcomes slip: run root-cause (data/model/process), apply decision triggers, and rebalance work
Guardrail (leading)	Override rate (by reason); exception backlog; decision evidence completeness (logging coverage); data quality SLA breaches; drift alerts	If guardrails breach: increase human review, tighten autonomy tier. pause automation, and escalate via the response playbook
Outcome (lagging)	Cost-to-serve / cost-to-income; CSAT/NPS; retention; fraud losses; early delinquency	Confirm benefits; decide whether to scale to adjacent journeys
Guardrail (lagging)	Findings; operational losses/incidents; time-to-remediate; confirmed bias/fairness issues	Re-baseline controls, update the response playbook, and report remediation status to governance

The 2x2 Executive Metric Lens

Use this 2×2 to build an executive dashboard: pair outcome KPIs (value delivered) with guardrail KPIs (control effectiveness) and track leading (early warning) and lagging (realized impact) indicators. Define decision triggers once per workflow (see Appendix B) so metrics drive action, not just reporting.

Table 2. The 2x2 Executive Metric Lens (Outcome/Guardrail x Leading/Lagging)

	Leading indicators (early warning)	Lagging indicators (realized impact)
Outcome KPIs (Value delivered)	• In-flight cycle time/ queue age • STP rate (by journey step) • Digital adoption/ active users	• Cost-to-serve/ cost-to-income • CSAT/NPS (surveyed) • Retention/ churn • Loss ratios (fraud loss rate, credit delinquency)
Guardrail KPIs (Control effectiveness)	• Override rate (by reason code) • Exception volume/ backlog • Decision evidence completeness (logging coverage) • Model drift/ data quality SLA breaches	• Confirmed control breaches/findings • Operational losses/ incident severity • Time-to-remediate breaches • Fairness/bias issues confirmed in review

Minimum viable dashboard: Start with one or two outcome KPIs and one or two guardrails per priority journey. Review at a cadence that matches the workflow (daily/weekly for critical controls). When signals deteriorate, intervene early (tighten autonomy, add review or pause automation).

- **Monitor autonomy where it matters:** Focus measurements on decision points (routing, approval, escalation). Keep the scorecard small and comparable across journeys.
- **Tailor by domain:** Use the domain table to select outcome metrics (experience, cost, speed) and the guardrails required by that business area.
- **Include a few enabling signals:** Track the tech and data indicators that predict business impact (availability, recovery time, security response, data quality, drift).
- **Start with priority journeys:** Select 2–3 high-impact journeys; define baseline, targets, and guardrails.
- **Set decision rights and decision triggers:** Define autonomy tiers, owners, monitoring cadence and stop-the-line triggers (override rates, drift, policy breaches).
- **Instrument and scale safely:** Implement logging and dashboards, measure lift vs. baselines, strengthen controls and scale only when outcomes improve and guardrails stay stable.

Importance to regulators: Supervisors are moving toward dashboards and continuous evidence of resilience, fairness and data quality. Expectations increasingly include metrics for auditability, model bias and override frequency as part of model-risk and operational-resilience oversight. Examples include the Federal Reserve's SR 11-7 guidance on model risk management (Board of Governors of the Federal Reserve System, 2011), the EU's Digital Operational Resilience Act (European Union, 2022), NIST's AI Risk Management Framework (NIST, 2023), and BCBS 239 principles on risk data aggregation and risk reporting (Basel Committee on Banking Supervision, 2013; 2023), which support metric integrity and executive reporting.

The Strategic Shift to AI-native, Agentic Operating Models

From periodic reporting to operating oversight: Move from retrospective KPIs to indicators that are updated frequently enough to manage the workflow (cadence depends on the risk and customer impact).

- **Continuous signals where feasible:** Supplement surveys and monthly views with interaction and journey telemetry (e.g., sentiment from calls/chats).
- **Decision triggers:** Define what happens when signals breach (route, increase review, pause automation, escalate). See Appendix B for metrics catalog examples and Appendix C for definitions.
- **Agentic workflows:** As AI moves from recommendations to execution, define an autonomy tier for each workflow and instrument it for operating-speed oversight.
- **Effectiveness:** cycle time, straight-through processing (STP), resolution rates.
- **Guardrails:** false positives/false negatives, override rate, decision evidence completeness.
- **Culture and organization:** AI-native metrics work best with clear end-to-end accountability. Give each journey team a concise scorecard and review it consistently.

Decision rights and tiers: Define an autonomy tier for each agentic workflow (assistive → semi-autonomous → higher autonomy) and pre-define what changes by tier:

- Override paths and roles.
- Confidence thresholds and stop-the-line triggers.
- Evidence requirements (logs, explanations, outputs) sufficient for audit and review.

Example: When moving a workflow from assistive to semi-autonomous, firms typically increase sampling and review SLAs, tighten stop-the-line thresholds, and require more complete decision logging (inputs, outputs, versioning, and rationale).

Operating at Scale without Sacrificing Control

At scale, AI-native metric design pairs performance gains with explicit control measures so leaders can see both improvement and risk in the same view:

- **Cycle time vs. quality:** If an agent reduces credit underwriting time, also monitor default rates, policy exceptions, and model usage limits to ensure speed does not increase risk.
- **False positives vs. true risk:** If a model reduces AML alert volume, track SAR conversion and other true-positive indicators to confirm real suspicious activity is still being detected. Pair fraud "detection rate" with fraud losses and customer friction to avoid shifting cost to customers through false flags.
- **Customer convenience vs. escalation:** If a chatbot resolves most inquiries end-to-end, monitor first-contact resolution and escalation-to-human rates to ensure complex issues are managed appropriately.

At scale, metrics should trigger interventions. When performance or controls drift, route work, tighten autonomy, add review or pause automation; and ensure evidence is captured for review.

Evolving Metrics by Business Domain (Highlights)

BFSI organizations are categorizing metrics along two dimensions: (1) Business domain – reflecting lines of business, and (2) technology and operational domains, which support all lines of business. The interplay between these categories highlights how AI is impacting both the front-end business outcomes and the back-end capabilities needed to sustain them.

Domain	Top outcome metrics (examples)	Top guardrails (examples)
Retail & commercial banking	Digital adoption; cycle time; STP; cost-to-serve; retention	Fair lending indicators; policy exceptions; drift; decision evidence completeness
Wealth management	AUM/net new assets; retention; advisor productivity; service experience	Suitability/KYC completeness; surveillance true-positive yield; review timeliness
Capital markets & investment banking	Execution quality; deal throughput; trade breaks; client coverage productivity	Market/liquidity limit breaches; surveillance quality; operational resilience incidents
Financial crime & compliance	Investigator throughput; time-to-file; backlog; customer friction reduction	False negatives + true-positive yield; QA defects; decision evidence completeness
Insurance	Claims cycle time; STP; loss/combined ratio; renewal/ persistency	Leakage/fraud guardrails; exception rates; drift/ quality of triage decisions

Put AI Native Metrics into Operating Rhythm

If you are scaling AI-enabled or agentic workflows, use this paper as a practical starting point for an executive-ready measurement system. A lightweight, 30–45-day sprint is often enough to create a minimum viable scorecard, clarify decision rights and establish evidence-ready guardrails.

1. **Select 2–3 priority journeys** (highest value × highest risk) and populate the executive dashboard with a small set of outcome and guardrail metrics.
2. **Baseline and instrument key signals** (cycle time, STP-by-step, override rate, evidence completeness, drift/quality alerts) so you can manage weekly, not quarterly.
3. **Define decision triggers** and autonomy tiers for each workflow (when to route, add review, tighten thresholds or pause automation).
4. **Assign accountable owners** and confirm the control partners (model risk, compliance, ops risk, SRE) who sign off on guardrails and response playbooks.
5. **Validate with primary evidence** (policy, logs, QA sampling, remediation records) to ensure the scorecard is audit and regulator-ready.

Synechron can facilitate an executive scorecard working session and help translate these templates into a governed operating rhythm tailored to your products, jurisdictions, and risk appetite.

Appendices

Appendix A. Domain Deep Dives

The table below summarizes the most important operating-model shifts driven by AI and agentic workflows, along with practical implications for BFSI institutions and their technology partners.

Retail and Commercial Banking Metrics

Key Business Priorities: Digital adoption and engagement, customer experience, retention and growth, credit quality and operating efficiency.

- **Digital adoption and active users:** Track the share of customers regularly using mobile/online channels, key drivers of cost-to-serve and convenience.

AI's role: *personalization, virtual assistants and journey analytics to lift engagement and conversion.*

- **Client experience (CSAT/NPS):** Leading indicators of retention and growth, increasingly complemented by always-on sentiment signals.

AI's role: *NLP turns calls/chats/complaints into service insights and prioritizes fixes.*

- **Retention & churn:** Track who stays and who leaves—often cheaper to protect than reacquire.

AI's role: *predict at-risk customers early and trigger targeted interventions to reduce attrition.*

- **Cross-sell & upsell:** Track products-per-customer and offer conversion to grow share of wallet.

AI's role: *"next best offer" recommendations that improve relevance and conversion.*

- **Credit risk & lending:** Track delinquency/NPLs, approval rates, and underwriting time—core portfolio health and efficiency signals.

AI's role: *improve risk prediction and speed decisioning, with monitoring for quality and fairness.*

- **Cost-to-income ratio:** Operating expenses as a share of operating income—a top executive efficiency KPI.

AI's role: *automation and analytics reduce manual effort and rework, improving the ratio over time.*

- **Digital banking in action at Bank of America** *(example; illustrative, not a universal benchmark):* Bank of America's public reporting illustrates what "AI and digital" metrics can look like at scale.

Comparability note: The metrics below are reported as stated by the source and may not be directly comparable across institutions due to differences in definitions, scope, product mix, and reporting methods.

  - **~30B digital client interactions (2025):** ~30B connections via digital logins and proactive alerts; **+14%** YoY *(Bank of America Newsroom, Mar. 10, 2026).*
  - **3.2B+ Erica interactions (since 2018 launch):** Cumulative client interactions with Erica since launch *(Bank of America Newsroom, Mar. 10, 2026).*
  - **Digital channel usage:** 81% consumer & small business households; 86% wealth management; 86% global banking clients engage digitally *(Bank of America Newsroom, Feb. 24, 2026, "Bank of America Fast Facts").*
  - **Digital satisfaction:** 86% of clients rate their digital experience **9/10+** *(Bank of America Newsroom, Feb. 24, 2026, "Bank of America Fast Facts").*
  - **Life Plan payoff (as of Dec 2022):** Clients who used Life Plan added **$55B+** to Bank of America accounts *(Bank of America Newsroom, Feb. 24, 2026, "Bank of America Fast Facts").*
- **Document processing** (example; illustrative, not a universal benchmark): JPMorgan Chase & Co.'s Annual Report 2016 (p. 49) describes COiN reviewing ~12,000 annual commercial-loan agreements in seconds—work previously estimated at ~360,000 hours/year *(secondary coverage: Bloomberg, Feb. 27, 2017)*
- **Industry estimate** *(illustrative scenario; not a universal benchmark):* Strategy& (PwC) estimates AI could improve bank cost-to-income by ~10–15 points over time, assuming strong controls prevent risk, compliance, and operational costs from offsetting savings (Strategy&, 2025).

Wealth Management Metrics

Key business priorities: Asset growth, retention, advisor productivity, and compliance.

- **AUM:** Total client assets with the firm—drives fee income and signals trust.

AI's role: *help advisors scale personalization and surface "money-in-motion" signals for timely outreach.*

- **Investment performance (alpha, Sharpe):** Track excess and risk-adjusted returns—persistent underperformance drives outflows.

AI's role: *accelerate research (news/filings/alt data) and run scenarios more frequently to support timely rebalancing.*

- **Client experience & retention (NPS, churn):** *Key growth levers as switching costs fall.*

AI's role: *use sentiment and behavior signals to trigger proactive service recovery and next-best actions.*

- **Wealth compliance (suitability, KYC, surveillance):** *High-scrutiny controls for fiduciary duty and financial crime.*

AI's role: *automate onboarding evidence and enhance monitoring to surface mis-selling, conflicts, and market-abuse risk earlier. IOSCO has highlighted these use cases and associated risks in AI and machine learning in asset management and related services (IOSCO, 2021).*

- **Controls to watch:** *onboarding turnaround time, file completeness, surveillance true-positive yield, and review timeliness.*
- **Digital wealth in action @ Morgan Stanley** *(example; illustrative, not a universal benchmark):* Public reporting provides a scale/context reference point for wealth and investment management metrics (use primary earnings materials for external publication).

Comparability note: Figures are reported as stated by the source and may not be directly comparable across institutions due to differing definitions (e.g., AUM vs. client assets), scope, and reporting methods.

- **Morgan Stanley (publicly reported, FY2024):** (i) **$7.9T** total client assets across Wealth and Investment Management; (ii) **$2,347B** fee-based client assets in Wealth Management; (iii) **$123.1B** fee-based asset flows; and (iv) **$251.7B** net new assets (Morgan Stanley, Jan. 16, 2025, "Morgan Stanley Fourth Quarter and Full Year 2024 Earnings Results").

Capital Markets and Investment Banking Metrics

Key Business Priorities: Market and liquidity resilience, trading performance and controls, deal pipeline conversion, and operational risk management.

- **Example (industry benchmark; illustrative):** JP Morgan Markets' The e-Trading Edit 2026: The results are in reports **955** institutional trader respondents across **13** asset classes and **59** global locations—useful context for "digital markets at scale" when tracking adoption of electronic workflows (J.P. Morgan Markets, 2026).
- **Example (firm scale context; not AI-specific; illustrative):** JPMorgan Chase & Co.'s April 7, 2025, press release publishing its 2024 annual report notes approximately **$4.0T** in total assets as of **Dec. 31, 2024**—useful context for the operational scale behind trading/risk controls (JPMorgan Chase & Co., 2025).

Comparability note: Institution-reported figures are included as contextual examples and may not be directly comparable across firms due to differences in definitions, consolidation scope, and reporting periods.

- **Recommended control-quality metric:** surveillance true-positive yield (% of alerts leading to a case) and time-to-triage for top-priority anomalies.
- **Trading & execution quality:** Track P&L, volumes, hit rate, and slippage—core profitability and client-trust measures.

AI's role: *optimize routing/execution and strengthen surveillance by prioritizing anomalies for review.*

- **Market risk & liquidity (VaR, stress, LCR):** Board-level resilience indicators under close supervision.

AI's role: *near-real-time aggregation/recalculation and anomaly detection for unusual exposure spikes.*

- **Deal pipeline & client coverage:** Track pipeline, win rates, and league tables—leading indicators of advisory revenue.

AI's role: *trigger analytics to focus coverage and GenAI to accelerate first drafts (measure time saved and downstream outcomes/quality).*

- **Operational & control metrics:** Track trade breaks/failed trades and responsiveness—small failures can scale into losses and regulatory exposure.

AI's role: *detect anomalies early and prioritize surveillance alerts; measure timeliness and true-positive yield (not just alert volume).*

Financial Crime and Compliance Metrics

Key Business Priorities: Detect illicit activity, reduce false alerts, meet filing timelines, and maintain decision evidence completeness.

- **AML alerts & false positives:** High false positives consume investigative capacity and can delay true-risk detection.

AI's role: *add context (history, peers, networks) to improve precision—track false positives, true-positive yield, and missed-case reviews together.*

- **SAR timeliness & quality:** Regulators expect on-time filing and complete narratives.

AI's role: *draft narratives from case data (with human accountability) and track time-to-file, QA defects/rework, and escalation rates.*

- **KYC onboarding & reviews:** Slow or inconsistent KYC creates customer friction and backlog risk.

AI's role: *automate verification and risk scoring to route cases; track turnaround time, file completeness, and exception/error rates. FATF's Guidance on digital identity provides a reference point for control considerations (FATF, 2020).*

- **Fraud detection & prevention:** Balance detection with customer friction—over-blocking erodes trust.

AI's role: *real-time behavioral/device scoring; track false positives/false declines alongside losses prevented and confirmed fraud capture.*

- **Regulatory reporting accuracy & timeliness:** Late or incorrect filings can trigger findings and remediation.

AI's role: *automate lineage/reconciliation and validation; track defect rate, rework, and time-to-submit.*

Insurance Metrics

Key Business Priorities: Claims experience, underwriting profitability, and risk resilience.

Publication note (credibility): This section avoids vendor-reported performance claims; when citing claims KPIs externally, prefer primary insurer disclosures (annual reports, earnings materials, statutory filings) and clearly define the scope (claim type, channel, complexity tier).

- **Loss ratio & combined ratio:** Core underwriting profitability metrics.

AI's role: *improve selection/pricing (e.g., telematics) and reduce leakage via fraud detection and automation; validate with profitability by product/segment over time.*

- **Claims (FNOL settlement):** Cycle time is a primary driver of trust and expense.

AI's role: *automate triage/intake and route complex or high-risk claims earlier; track STP rate, cycle time, and post-claim satisfaction.*

- **Renewal / persistency:** Retention stabilizes the book and drives lifetime value.

AI's role: *predict lapse risk and link outcomes to experience drivers (service, claims satisfaction, pricing) to target fixes.*

- **Capital adequacy / solvency ratios:** Key indicators for ratings and supervisory confidence.

AI's role: *accelerate scenario simulation and monitor underlying drivers (market, credit, underwriting risk) with alerts on material shifts.*

Case Example – AI in Fraud and Underwriting:

Illustrative outcomes from AI-enabled operations (examples only; not a universal benchmark—validate figures and assumptions against primary sources before external publication):

- **Fraud prevention:** Network-scale pattern detection can materially reduce fraud losses when paired with precision metrics (false declines) and confirmed-loss tracking.
- **Underwriting efficiency:** AI-assisted underwriting can reduce time-to-quote/issue; track cycle time and post-bind loss performance to confirm quality is maintained.
-
- **Customer experience:** Faster claims and fewer handoffs can lift satisfaction; monitor CSAT/NPS alongside fraud referrals and leakage to ensure speed does not increase risk.

Appendix B. Metrics Catalog (starter set)

Use this starter catalog to standardize definitions, assign owners, and define **decision triggers** before scaling AI-enabled or agentic workflows. Tailor targets by product, segment, and risk appetite; pair each outcome metric with at least one guardrail metric and define the **response playbook** when a trigger is breached.

Table 4. Metrics Catalog (Starter Set)

Metric	Type	Definition (example)	Typical cadence	Owner / primary data source
Bias / fairness indicators	Data/Model Health (Guardrail)	Disparity metrics across protected classes where applicable; define thresholds and review workflow.	Monthly / quarterly	Model risk + compliance; model monitoring + fair lending analytics
Change failure rate and MTTR	Operational (DevOps)	% of deployments causing incidents + time to restore service; tie back to customer-impact minutes.	Weekly	SRE/Engineering; CI/CD + incident system
Cost-to-serve (per customer / per case)	Outcome (Cost)	Total operating cost allocated per completed case/journey; monitor before/after automation.	Monthly	Finance + journey owner; ERP + activity-based costing inputs
Customer satisfaction (CSAT) / NPS	Outcome (Experience)	Customer-reported satisfaction; supplement with always-on sentiment from interactions.	Weekly / monthly	CX owner; survey platform + contact center analytics
Cycle time (such as loan decision, KYC, claim)	Outcome (Speed)	Median time from intake to decision/closure; segment by channel and risk tier.	Daily / weekly	Ops owner; process telemetry + timestamped events
Data quality (accuracy/ completeness/ timeliness)	Data/Model Health	% of critical data elements meeting checks; include data freshness SLAs.	Daily / weekly	Data owner; data observability + data warehouse/lake
Decision evidence completeness	Guardrail (Auditability)	% of decisions with required logs/evidence captured (inputs, outputs, versioning, reviewer, rationale).	Weekly	Risk/controls; logging platform + GRC tooling
Default / delinquency rate (credit portfolio)	Outcome (Risk-adjusted)	Early delinquency (such as 30/60/90 DPD) for originated loans; cohort by model version and policy tier.	Monthly	Credit risk; core banking + risk data mart
Digital adoption / active users	Outcome (Adoption)	% of customers active in digital channels; track feature adoption for AI assistants/copilots.	Weekly / monthly	Digital product owner; channel analytics
False-negative rate (miss rate)	Guardrail (Control quality)	Estimated missed suspicious/fraud cases based on sampling, backtesting, and downstream discoveries.	Weekly / monthly	Compliance/Fraud + QA; audit samples + model validation outputs
False-positive rate (alerts, flags, blocks)	Guardrail (Control quality)	% of alerts/blocks ultimately cleared as non-issues; monitor alongside true-positive yield.	Daily / weekly	Compliance/Fraud; monitoring platform + investigator outcomes
Fraud loss rate	Outcome (Loss)	Confirmed fraud losses as % of volume or $ exposure; segment by channel and product.	Weekly / monthly	Fraud; case system + chargeback/ledger data
Latency and error rate (critical AI services)	Operational	P95/P99 latency and % errors for inference, orchestration, and dependencies; track by model/version.	Real-time / daily	Platform/SRE; observability stack (APM/logs)
Lineage coverage	Data/Model Health	% of key metrics/models with end-to-end lineage documented and kept current automatically.	Monthly	Data governance; catalog/ lineage tools
Model drift incidents	Data/Model Health	Count/rate of drift threshold breaches; track time-to-detect and time-to-mitigate per model.	Weekly	Model owner; MLOps monitoring + validation logs
Model inventory coverage	Governance	% of production models (including GenAI/agent components) registered with owner, purpose, tier, and controls.	Monthly	Model risk; model registry + GRC tooling
Override rate (human overrule / exception)	Guardrail (Governance)	% of agent/model recommendations overridden by humans (or policy exceptions granted); segment by reason code.	Daily / weekly	Journey owner + risk; decision logs + case notes
Security detection and response (MTTD/MTTR)	Operational (Security)	Time to detect and contain security events; segment by severity and asset class.	Weekly / monthly	Security; SIEM/SOAR + incident system
Straight-through processing (STP) rate	Outcome (Efficiency)	% of cases/transactions completed end-to-end without manual touch.	Daily / weekly	Journey owner; workflow system + case management
Time-to-remediate control breaches	Guardrail (Resilience)	Median time from breach detection to fix/closure; include repeat-breach rate.	Weekly / monthly	Ops + controls; incident management system

Appendix C. Terminology and definitions

The definitions below reflect how terms are used in this paper and are intended to support consistent executive and stakeholder interpretation.

How repeated terms are used in this paper: For readability, the main body uses shorthand. Operating-speed oversight means measurement on a cadence fast enough to intervene in the workflow. Decision triggers / playbooks means pre-agreed responses when metrics breach (route work, add review, tighten autonomy, pause automation, escalate). Decision evidence means the logs and artifacts needed to reproduce and review material decisions.

Table 5. Terminology and Definitions (as used in this paper)

Term	Definition (as used here)
Agentic workflow	A workflow where AI performs multi-step tasks and decisions with defined autonomy; humans supervise through exception handling, review, and override paths.
AI-native metrics	A KPI system designed for AI-enabled and agentic operating models, combining outcome measures with guardrails and near-real-time telemetry so metrics can trigger interventions—not just reporting.
Audit evidence completeness	The degree to which each material decision has sufficient records to support review (inputs, outputs, model/ workflow version, rationale, reviewer/approver where applicable).
Autonomy tier	A classification of how much decision and execution authority an agentic workflow has (such as assistive → semi-autonomous → higher autonomy), which determines review requirements, thresholds, and evidence standards.
False positives / false negatives	Incorrect alerts/flags (false positives) vs. missed true-risk events (false negatives). In controls, these should be monitored together with true-positive yield to avoid reducing alerts at the cost of missed risk.
Governed autonomy	Scaling automation and agent execution within explicit boundaries (policy, risk appetite, controls), with continuous evidence and decision rights defined by tier.
Guardrail KPIs	Metrics that demonstrate control effectiveness and governance quality as autonomy increases (e.g., override rate, false positive/false negative rates, policy exceptions, audit evidence completeness, time-to-remediate).
Lineage coverage	The extent to which key metrics, reports, and models can be traced end-to-end to source data and transformations to support auditability and root-cause analysis.
Model drift	A statistically meaningful change in input data patterns or model performance over time that can degrade decision quality and increase risk; drift incidents are threshold breaches that require review or mitigation.
MTTD / MTTR	Mean Time to Detect and Mean Time to Restore/Recover—operational resilience measures for how quickly issues are found and services are returned to normal.
Operating-speed oversight	Monitoring and managing performance and controls on a cadence fast enough to intervene while decisions are being made (near real time for critical workflows), rather than only via periodic reporting.
Outcome KPIs	Metrics that reflect business and customer value delivered, such as cycle time, STP rate, cost-to-serve, CSAT/ NPS, retention and losses.
Override rate	The percentage of agent/model recommendations or actions that are overruled by a human (or forced into an exception path), often segmented by reason codes and tier.
Straight-through processing (STP) rate	The percentage of cases/transactions completed end-to-end without manual intervention.
Telemetry	The event, log, and performance data captured from workflows and AI systems (inputs, outputs, confidence, versions, latency, errors) used to measure outcomes, guardrails, and operational health.

Appendix D. Longer-term roadmap (2026–2030)

2026–2030 roadmap (summary): Key enablers for a more autonomous, continuously governed operating model.

Invest in data and platform readiness –
many institutions are still modernizing core systems and data lakes to be real-time and "AI-friendly."

Strengthen governance –
treat AI governance as an enabler of speed, not a hindrance. Those who embed controls early (from design stage) will scale faster.

Start small, learn fast –
pilot agentic processes in sandbox environments, measure, refine, then expand. Avoid big bang: use a portfolio of quick experiments with clear metric goals.

Collaborate externally –
share best practices through industry groups (e.g., BIS, IOSCO working groups on AI), and leverage guidance like NIST's AI Risk Management Framework or Deloitte's The measured leap: AI agent observability to design robust metric systems.

Sources

Basel Committee on Banking Supervision. (2013). Principles for effective risk data aggregation and risk reporting (BCBS 239). (accessed 2026-04-13).

Basel Committee on Banking Supervision. (2023). Progress in adopting the principles for effective risk data aggregation and risk reporting (BCBS 239). (accessed 2026-04-13).

Board of Governors of the Federal Reserve System. (2011). SR 11-7: Guidance on model risk management. (accessed 2026-04-13).

European Union. (2022). Digital Operational Resilience Act (DORA) (Regulation (EU) 2022/2554). (accessed 2026-04-13).

Financial Action Task Force. (2020). Guidance on digital identity. (accessed 2026-04-13).

International Organization of Securities Commissions. (2021). AI and machine learning in asset management and related services. (accessed 2026-04-13).

National Institute of Standards and Technology. (2023). AI Risk Management Framework (AI RMF 1.0). (accessed 2026-04-13).

Industry research and practitioner reports

Strategy& (PwC). (2025, Oct. 16). The future of banking: How AI is reshaping the industry. (accessed 2026-04-13).

Deloitte. (2026, Feb. 18). The measured leap: AI agent observability — Appraising multiagent systems impact with agent operations. (accessed 2026-04-13).

Company examples and secondary reporting

Bank of America Newsroom. (2026, Feb. 24). Bank of America Fast Facts. (accessed 2026-04-13).

Bank of America Newsroom. (2026, Mar. 10). BofA AI and Digital Innovations Fuel 30 Billion Client Interactions. (accessed 2026-04-13).

Bloomberg. (2017, Feb. 27). JPMorgan software does in seconds what took lawyers 360,000 hours. Source note: COiN statistics are cited from JPMorgan Chase & Co. (2017), Annual Report 2016, p. 49; Bloomberg is included as secondary corroboration. (accessed 2026-04-13).

J.P. Morgan Markets. (2026). The e-Trading Edit 2026: The results are in (E-Trading Survey Report page; 10th annual e-Trading survey; 955 institutional trader respondents; 13 asset classes; 59 global locations). (accessed 2026-04-13).

JPMorgan Chase & Co. (2017). Annual Report 2016 (official Investor Relations PDF), p. 49. (accessed 2026-04-13).

JPMorgan Chase & Co. (2025, Apr. 7). JPMorganChase Publishes 2024 Annual Report, Including Chairman & CEO Letter to Shareholders (press release; notes $4.0T total assets as of Dec. 31, 2024). (accessed 2026-04-13).

Morgan Stanley. (2025, Jan. 16). Morgan Stanley Fourth Quarter and Full Year 2024 Earnings Results. (accessed 2026-04-13).

Author

Frances Cruz is Head of Consulting, Americas at Synechron. She advises financial services leaders on operating-model transformation across digital, data and AI, connecting measurable business outcomes with practical governance, risk and control frameworks needed to scale safely.

Disclaimer

This white paper is provided for general informational purposes only and reflects the author's views as of the publication date. It does not constitute legal, regulatory, tax, investment, or other professional advice, and it should not be relied upon as a substitute for consultation with qualified advisors. References to laws, regulations, supervisory expectations, and industry practices are generalized and may vary over time and by jurisdiction. Any examples, use cases, metrics, figures, or third-party references are illustrative; readers should validate assumptions and results using primary sources and their own data, controls, and risk appetite. The content is provided "as is," without warranties of any kind, and Synechron disclaims liability for actions taken based on this material.

No client endorsement or affiliation is implied by any references to third-party organizations, institutions, products, or services.