Whitepaper

AI Risk and Governance Index

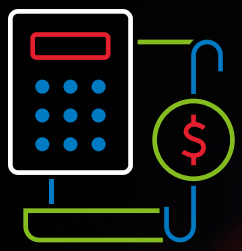
US Banking

H1 2026

Wolters Kluwer

wolterskluwer.com

Executive introduction



2026 marks the year AI transitions from banking's most-discussed ambition to its most scrutinized operational reality. Institutions have spent three years piloting predictive models, deploying GenAI copilots, and embedding detection algorithms into fraud and AML workflows. The conversation has shifted — not from whether to adopt AI, but how to govern it, how to protect customers through it, and whether the organizational infrastructure exists to sustain it under continuous regulatory supervision.

This report draws on survey responses from 230 US banking professionals spanning community banks under $1 billion AUM to large institutions above $50 billion. The panel is senior: 30 percent hold Director or Sr. Director titles, with C-Suite and VP-level respondents comprising an additional 31 percent. Functional representation covers credit risk, compliance, model risk management, regulatory reporting, fair lending, and AI architecture — with the largest single role cohorts being Head of Credit Risk/Underwriting (9.57 percent) and Head of Regulatory Reporting/Data Engineering (9.13 percent)). This is not a technology adoption survey — it is a risk and governance survey from the practitioners who bear institutional accountability for getting AI right.

This report's peer benchmark data helps readers calibrate their own AI risk and governance posture against institutions facing comparable regulatory, operational, and reputational pressures. In an environment where internal progress can be difficult to contextualize, benchmarks provide an external reference point for distinguishing true maturity from local optimization. Used appropriately, they are not scorecards, but decision aids — highlighting where governance investments are converging across the industry, where divergence may introduce supervisory risk, and where leadership attention is most likely to yield disproportionate risk reduction.

The findings organize into three thematic domains: the integrity of the models and data that power AI systems; the fairness and consumer protection implications of how those systems make decisions; and the human, organizational, and vendor readiness required to operate and remediate AI safely at scale. Across all three, a consistent signal emerges — deployment velocity has outpaced governance maturity, and the gap is widening in the functions where the consequences of failure are highest.

Section 1: Model and data governance

Executive overview

The foundational premise of trusted AI is that a model is only as reliable as the governance architecture surrounding it — the validation frameworks that certify its performance, the data pipelines that feed it, the monitoring systems that track its behavior over time, and the lineage documentation that makes it auditable on demand. What emerges from this section is a picture of a sector that has built meaningful AI capability on a governance foundation that has not kept pace.

Model governance is not a compliance function that sits downstream of AI deployment — it is the prerequisite that determines whether deployment can scale at all. Institutions that validate models at launch but lack the infrastructure to monitor performance continuously, detect drift, or trace individual decisions through a documented lineage are not governing their AI. They are gambling on stability in environments that are structurally dynamic. Customer behavior shifts, macroeconomic conditions change, and adversarial actors adapt — and all three forces can silently degrade a model's accuracy and fairness without triggering any internal alert.

Data risk is the second, inseparable dimension of this challenge. A model that is governed but fed dirty, synthetic, or inadequately traced data produces outputs that are defensible in form but unreliable in substance. The emergence of synthetic data generation as a primary AI development tool has introduced a new class of risk that practitioners are only beginning to quantify: models trained on synthetic distributions that diverge from actual customer behavior in ways that may not surface until a downstream decision causes harm. The survey findings across this theme reflect a sector that understands these risks clearly — and is still building the infrastructure to address them with the rigor the moment requires.

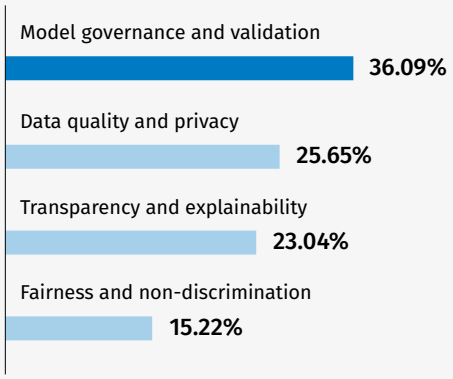
Governance tops the risk register

Model governance and validation leads decisively at 36.09 percent — nearly double the share who named **fairness and non-discrimination** (15.22 percent). The gap is significant: governance mechanics, not bias or explainability, is the primary constraint practitioners identify. This does not diminish the importance of fairness; it reflects that governance and data quality must be resolved first. Institutions cannot run meaningful bias testing on models they cannot trace, and cannot explain decisions made by systems they cannot consistently monitor. **Data quality and privacy** follows closely at 25.65 percent, reinforcing that the data foundation underneath governance is a near-equal concern.

Advisory signal: Institutions whose AI investment thesis is built around explainability and fairness tooling — without parallel investment in model validation infrastructure — are sequencing their priorities incorrectly. Governance is the foundation on which every other trusted AI capability is built.

> ***Model governance and validation*** *is the top barrier to safely scaling AI — cited by more than 1 in 3 respondents across 230 banking professionals, and more than double the share who named* ***fairness and non-discrimination****."*

Survey question: Which trusted AI risk factor represents the greatest challenge to safely scaling AI across your banking functions?



Percentage of respondents

More than 1 in 3 cite governance — not bias or explainability — as the primary scaling barrier.

Source: Trusted AI Survey, April 2026 | 230 respondents

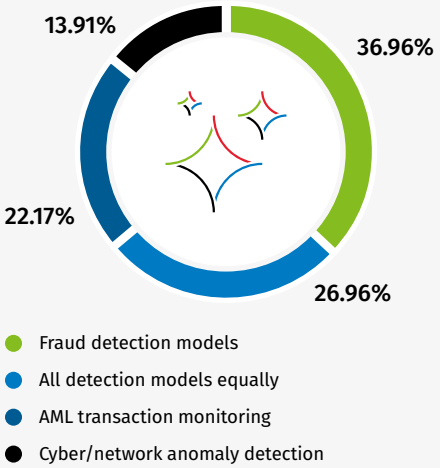
Fraud models live on the edge

Fraud models operate in adversarial environments where threat actors continuously adapt to circumvent detection logic. A model accurate today may be materially degraded within months — without any internal alert — in an organization without live performance monitoring. The 26.96 percent who flagged all detection models equally signals that for a meaningful cohort, drift and governance concern is not confined to one system type. Together, these two responses account for nearly two-thirds of the panel.

Advisory signal: Fraud detection models require a monitoring cadence that matches the velocity of the threat environment they are designed to detect. Annual validation cycles are structurally misaligned with adversarial drift timelines. Model owners should establish automated performance threshold alerts and define explicit revalidation triggers tied to detection rate degradation — not calendar schedules.

Fraud detection models *are the highest-risk detection category — outpacing AML monitoring and anomaly detection — driven by adversarial dynamics that make model drift a near-constant operational condition, not a periodic maintenance concern.*

Survey question: Which detection AI area presents the highest risk of model drift, false positives, or governance challenges?



36.96%
of fraud detection models flagged as highest drift and governance risk.

Source: Trusted AI Survey, April 2026 | 230 respondents

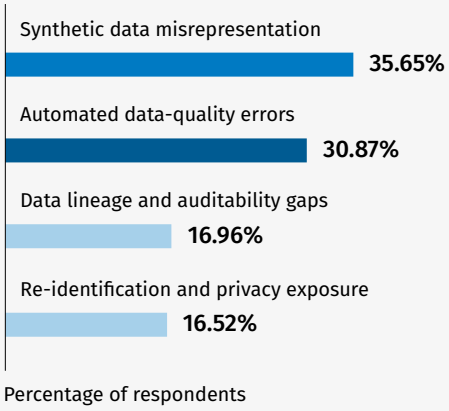
Synthetic data requires deeper governance

The tight clustering of the top two responses — synthetic misrepresentation and automated quality errors, together accounting for 66.52 percent of responses — reflects a shared concern about what happens when AI is used to generate or clean the data that other AI systems then train on. Error amplification across layered systems is a non-trivial risk that existing data governance frameworks were not designed to anticipate. The relatively lower scores for lineage gaps and re-identification likely reflect that those risks are more familiar and partially addressed; the synthetic data risks are newer and less mapped.

Advisory signal: Synthetic data governance must be treated as a distinct domain from traditional data quality management. Validation criteria, audit documentation, and model testing protocols for synthetic training data require purpose-built standards — not extensions of existing data lineage frameworks.

Synthetic data misrepresentation *tops the data risk register — as banks scale AI training pipelines, what feeds the model has become as consequential as the model itself.*

Survey question: Which data AI risk is most concerning as your institution expands use of data-quality automation and synthetic data generation?



66.52%
of responses concentrated in two emerging data failure modes

Source: Trusted AI Survey, April 2026 | 230 respondents

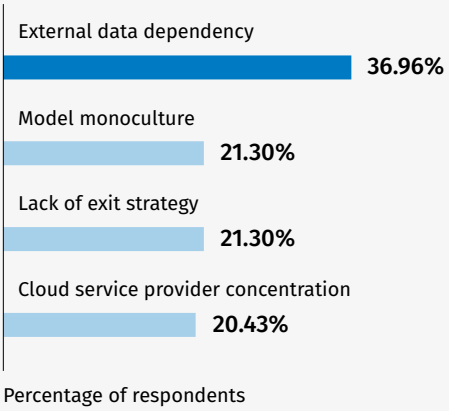
AI risk requires a portfolio-level dependency view

The near-uniform distribution across the three non-leading responses is itself a finding: systemic AI risk is multi-directional. Institutions are exposed to external data failure, homogenized model logic, vendor lock-in with no credible off-ramp, and cloud concentration risk — simultaneously, not in isolation. **Model monoculture** and **lack of exit strategy** tied exactly at 21.30 percent, with **cloud concentration** close behind at 20.43 percent — a three-way near-tie that underscores the distributed nature of structural AI risk. **External data dependency** is particularly consequential because it is largely invisible in traditional model risk frameworks, which tend to focus on internal model logic rather than upstream dependency chains.

Advisory signal: Systemic AI risk requires a portfolio-level view mapping all four dependency types — data, model, vendor, and cloud. Institutions without a dependency inventory cannot manage the concentration they cannot see.

External data dependency *is the single largest systemic AI risk — for most institutions, the integrity of their AI operations is anchored in data sources they do not fully control.*

Survey question: Which structural dependency introduces the highest level of systemic risk to your AI operations?



Model monoculture, lack of exit strategy, and cloud concentration within 0.87 pts of each other — structural AI risk is multi-directional.

Source: Trusted AI Survey, April 2026 | 230 respondents

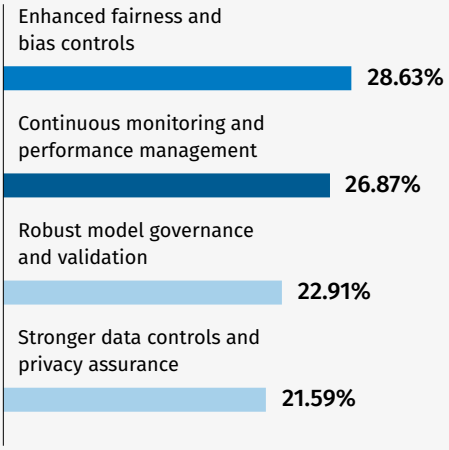
Control investment should be distributed across all domains

In an important shift from earlier data, **enhanced fairness and bias controls** leads at 28.63 percent, with **continuous monitoring and performance management** close behind at 26.87 percent — a gap of only 1.76 percentage points. The near-four-way distribution of responses across all categories (spanning just 7.04 percentage points from top to bottom) reflects a sector that does not believe control gaps are concentrated in one domain. Rather, practitioners see the urgency as broadly distributed, with fairness and monitoring competing at the top. This result reinforces the consumer protection findings in Section 2: as AI reaches more customer-facing functions, fairness control investment is being pulled forward as a first-tier priority.

Advisory signal: The near-parity across all four control needs is a portfolio investment signal, not a single-priority signal. Institutions that concentrate investment exclusively in monitoring or governance without concurrent fairness and data control investment will find their governance frameworks increasingly misaligned with examination priorities as regulators focus on AI consumer harm.

Enhanced fairness and bias controls *has emerged as the most urgently needed AI control — edging out* ***continuous monitoring*** *in a tightly contested result, and signaling a sector-level shift in where practitioners believe the most immediate gap lies.*

Survey question: Across all banking functions, where is the most urgent need to strengthen controls to reduce AI-introduced risk?



+1.76 pts separating the top two — control urgency is broadly distributed across all four domains.

Source: Trusted AI Survey, April 2026 | 230 respondents

Section 2: Consumer protection and fairness

Executive overview

The deployment of AI into customer-facing banking functions is not a technology initiative with fairness implications — it is a consumer rights issue with technology infrastructure. When an AI system determines whether a customer qualifies for a loan modification, how aggressively they are contacted about a past-due balance, or what settlement offer is surfaced during a collections interaction, the outcomes are subject to the same regulatory and ethical standards that govern any institutional decision affecting a customer's financial wellbeing.

Collections and recovery dominates the harm and exposure findings across multiple questions — not because it is the most technically complex AI deployment context, but because it is the one where AI-driven decisions interact most directly with financially vulnerable customers who have the fewest mechanisms for recourse. An underwriting decision that produces an adverse action notice, however imperfect, comes with a regulatory disclosure framework. A collections AI that applies inconsistent treatment, pursues contact at algorithmically optimized frequencies, or structures settlement offers that differ across customer segments based on predicted behavioral elasticity operates in a comparably less structured consumer-protection environment — and practitioners appear to understand that gap acutely.

The emergence of Agentic AI as a distinct risk category within this theme signals an important evolution in the conversation. Banks are no longer only concerned about static prediction model bias or single-point decision explainability. They are now grappling with AI systems that take sequences of autonomous actions across multi-step workflows, where human review may occur only at the beginning and end of a process — if at all. As agentic systems move from pilot to production in lending and collections workflows, the design of human-in-the-loop controls is not an implementation detail. It is the primary determinant of whether those systems are consumer-safe.

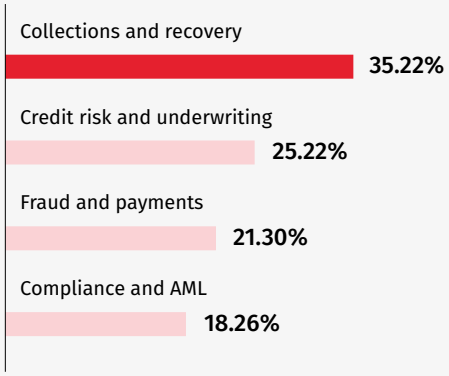
Collections is an emerging AI risk zone

Collections sits at the intersection of financial distress, behavioral AI, and limited regulatory disclosure requirements — making it a high-risk environment for AI-driven harm. **Credit risk and underwriting**, by contrast, operates within a more developed adverse action and fair lending framework that, while imperfect, provides a degree of customer recourse that collections lacks. The relatively even distribution across the three lower-ranked functions suggests that AI risk is perceived as diffuse across the customer lifecycle, not confined to origination.

Advisory signal: Every institution deploying AI in collections workflows — for contact strategy optimization, settlement offer generation, or payment plan eligibility — should assess whether its current governance, testing, and human oversight design would survive examination scrutiny. The question is not whether regulators will focus here; it is whether the institution will be ready when they do.

Collections and recovery *is the banking function where AI poses the greatest risk of customer harm — outpacing* ***credit risk and underwriting*** *by more than 10 percentage points.*

Survey question: In which banking function does AI introduce the highest risk of customer harm or regulatory exposure?



Percentage of respondents

Collections and recovery outpaces credit risk and underwriting by 10 percentage points.

Source: Trusted AI Survey, April 2026 | 230 respondents

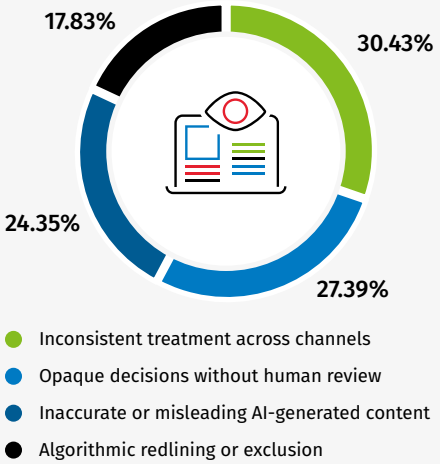
Test channel consistency, not just model outputs

The leading response is operationally specific in a way the others are not: inconsistent channel treatment is a systems integration and governance problem that exists independently of whether any individual model is biased. An institution whose AI surfaces different outcomes for equivalent customers across digital, telephonic, and in-person channels has a consumer protection problem that model fairness testing alone will not resolve. The tight clustering of the top three responses — accounting for 82.17 percent collectively — signals mature, practice-level understanding of where AI consumer harm actually materializes.

Advisory signal: AI fairness testing scope must extend across channel delivery, not just model outputs in isolation. A fair model deployed inconsistently across touchpoints produces unfair outcomes. Channel consistency testing must be a formal component of consumer protection governance.

Inconsistent treatment across channels *is the consumer-protection risk practitioners most anticipate — reframing AI fairness from a model bias problem into an operational consistency problem with direct fair lending implications."*

Survey question: Which consumer-protection risk is most likely to emerge as AI adoption expands?



82.17%
of responses named one of three specific, operationally actionable consumer risks.

Source: Trusted AI Survey, April 2026 | 230 respondents

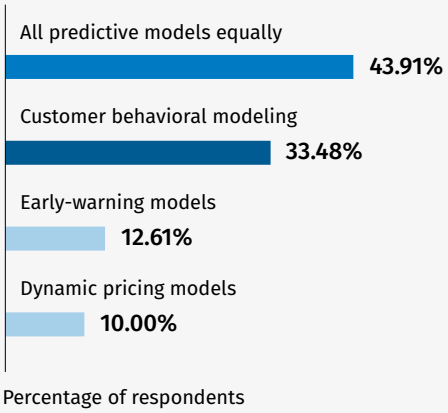
Fairness risk does not have a clear owner

The dominance of "all equally" reflects either a considered view that fairness risk is systemic to predictive AI in banking, or that institutions lack sufficient model inventory visibility to differentiate. Both interpretations have governance implications. **Customer behavioral modeling** as the strongest specific response at 33.48 percent reflects the risk that models designed to predict and optimize customer behavior can encode demographic or socioeconomic proxies that are technically neutral but practically discriminatory. **Dynamic pricing models**, despite public visibility, ranked lowest — likely reflecting limited current deployment in regulated banking relative to underwriting and behavioral scoring.

Advisory signal: The "all equally" plurality is a governance signal. If practitioners cannot identify which predictive AI capability poses the greatest fairness risk, their institutions likely lack sufficient model inventory visibility to make that determination. Comprehensive model cataloging is a prerequisite for risk-stratified governance.

A 43.91% plurality cannot isolate predictive AI fairness ***risk*** *to a single model type — signaling either a systemic risk view or a model inventory visibility gap that governance frameworks must address.*

Survey question: Which predictive AI capability introduces the most significant governance and fairness risk?



77.39%
of responses in two answers — systemic risk awareness or inventory visibility gaps.

Source: Trusted AI Survey, April 2026 | 230 respondents

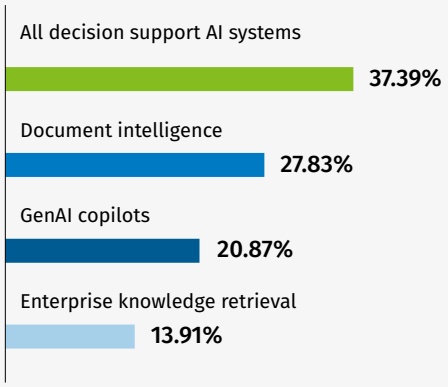
Document intelligence is the quiet risk

Document intelligence leads specific tool types at 27.83 percent, edging out **GenAI copilots** (20.87 percent) by nearly 7 percentage points. **Document intelligence** systems operate in regulated workflows — loan files, compliance records, examination documentation — where extraction errors carry direct legal and regulatory consequences. The gap between **document intelligence** and **enterprise knowledge retrieval** (13.91 percent) has also widened, suggesting practitioners are more clearly distinguishing where within the decision support stack the highest risk resides. The leading "all equally" response (37.39 percent) still reflects broad category concern, but the specific tool distribution is sharpening.

Advisory signal: Institutions whose AI risk governance singles out **GenAI copilots** as the primary decision support concern may be under-governing **document intelligence** systems operating with equivalent — or greater — consequence in regulated workflows. The full decision support AI inventory requires explicit risk assessment — not the assumption that the most visible tool type is the most risky.

Document intelligence *ranks above* ***GenAI copilots*** *as the highest-risk specific decision support AI tool — a signal that regulated document workflows carry AI risk that is frequently under-governed relative to the attention paid to generative AI.*

Survey question: Which decision support AI capability poses the greatest operational or compliance risk?



Percentage of respondents

Document intelligence leads specific tool types — 6.96 pts ahead of GenAI copilots.

Source: Trusted AI Survey, April 2026 | 230 respondents

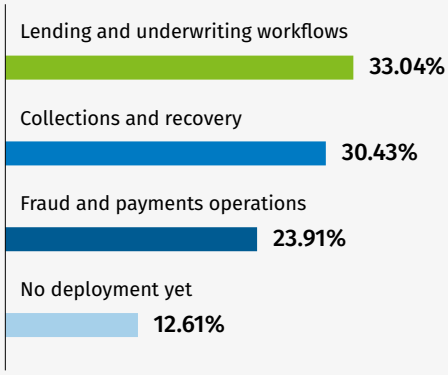
Autonomous AI, vulnerable customers

The 12.61 percent who report no current agentic AI deployment is notable: the vast majority of institutions are already deploying or planning agentic AI in high-stakes customer-facing functions. The convergence of agentic AI risk with the two highest-harm functions identified elsewhere in the survey is the most operationally urgent correlation in the dataset. Automation without human review in collections is not an abstract governance concern — it is a live customer exposure that regulators are beginning to examine with increasing specificity.

Advisory signal: Human-in-the-loop controls for agentic AI should not be designed as oversight of final outputs. They must be embedded as checkpoints within multi-step workflows — particularly at decision nodes that determine contact strategy, settlement eligibility, and account treatment in collections and underwriting environments.

Lending and underwriting, *along with* ***collections and recovery,*** *account for 63.48% of agentic AI risk concern. These are the same two functions that dominate the consumer harm findings, confirming that autonomous AI and customer vulnerability are converging in the same workflows.*

Survey question: Where does agentic AI introduce the greatest risk due to automation without sufficient human-in-the-loop controls?



Percentage of respondents

63.48%
of concern concentrated in the two highest-harm customer-facing functions identified elsewhere in this survey.

Source: Trusted AI Survey, April 2026 | 230 respondents

Section 3: Operational readiness and human risk

Executive overview

Governance frameworks and fairness controls are only as effective as the organizational infrastructure capable of operating them — and for a significant share of US banking institutions, that infrastructure is underdeveloped relative to the AI systems it is meant to oversee. This section examines the operational and human dimensions of AI risk: whether institutions can respond when models fail, whether third-party vendor relationships are managed with appropriate rigor, and whether the human factors that shape AI safety — incentives, cognitive biases, talent gaps, and rogue deployments — are receiving governance attention commensurate with their influence.

The incident response findings are among the most consequential in the survey. **Regulatory reporting for AI failures** and **model kill-switch protocols** are the two areas where institutions feel least prepared — and they are precisely the two capabilities that regulators will demand first when an AI system causes harm, attracts supervisory attention, or generates a material adverse event. Banks that have not rehearsed AI incident response workflows are building a remediation liability that compounds with every quarter of continued deployment. The question is not whether an AI incident will occur; it is whether the institution will be able to respond with the speed, accuracy, and accountability that regulators and customers will expect.

The human risk findings in this survey cycle carry a notable shift: **automation bias** has overtaken **misalignment of incentives** as the top human-centric AI risk.
This matters. Automation Bias — the tendency for human reviewers to defer to algorithmic outputs rather than exercise independent judgment — cannot be addressed through model validation, documentation, or training alone.
It requires governance design that creates genuine friction at review steps, explicit accountability for override decisions, and performance frameworks that reward decision quality alongside decision speed. Safe AI is not only a technical architecture — it is an institutional culture with deliberate behavioral guardrails.

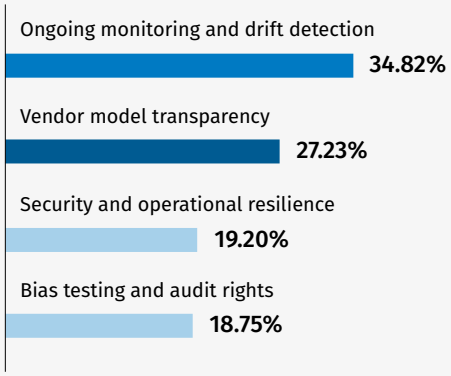
Vendor oversight demands constant surveillance

Ongoing monitoring and drift detection leads at 34.82 percent, with vendor model transparency second at 27.23 percent. Notably, security and operational resilience (19.20 percent) has edged ahead of bias testing & audit rights (18.75 percent) in this wave — a small but meaningful shift that may reflect growing operational risk awareness as agentic and embedded AI vendor deployments expand.
The near-parity of the two lower-ranked responses signals that vendor AI risk remains genuinely multi-dimensional. No institution is managing its third-party AI relationships well across all four vectors simultaneously, and the spread of concern reflects that reality.

Advisory signal: Vendor AI contracts that do not include explicit performance monitoring obligations, drift notification requirements, and defined audit rights are not fit for a continuous-supervision environment. Contract renegotiation cycles should be used to embed these requirements proactively — before examination pressure forces a reactive posture.

Ongoing monitoring and drift detection *is the hardest third-party AI vendor risk to manage — confirming that the live oversight gap in external AI relationships mirrors, and compounds, the gap in internal ones.*

Survey question: Which third-party AI vendor risk is most difficult for your institution to manage?



Security and operational resilience, along with bias testing and audit rights, are within 0.45 pts — vendor risk spans the full governance spectrum.

Source: Trusted AI Survey, April 2026 | 230 respondents

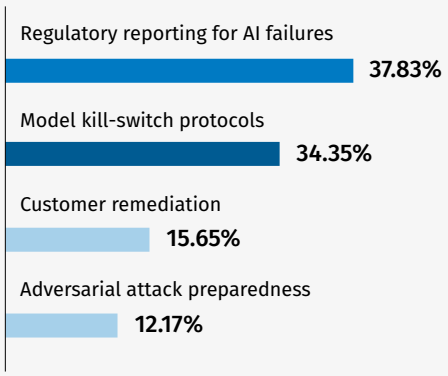
Institutions unprepared to stop or report a failure

The combined 72.17 percent concentration in the top two responses is the starkest readiness finding in the survey. **Regulatory reporting** and **kill-switch protocols** are not esoteric capabilities — they are the minimum viable requirements for managing an AI incident in a regulated environment. An institution that cannot accurately report an AI failure to regulators, and cannot stop a failing model from continuing to operate, is not equipped for the incident environment it is already living in. The relatively lower scores for **customer remediation** and **adversarial attack preparedness** likely reflect that these risks feel more distant — not that the institution is materially more prepared.

Advisory signal: Performance management frameworks that reward decision velocity, volume, or throughput without explicit weighting of decision quality and governance compliance are systematically building Automation Bias into AI-augmented workflows. Incentive redesign is a governance intervention — not an HR one.

*More than 7 in 10 respondents identified either **regulatory reporting** or **kill-switch protocols** as their institution's weakest incident response capability — the two functions regulators will demand first when AI systems fail.*

Survey question: In which area of AI incident response is your institution least prepared?



Percentage of respondents

72.17%
of respondents concentrated their least-prepared response in regulatory reporting or kill-switch protocols.

Source: Trusted AI Survey, April 2026 | 230 respondents

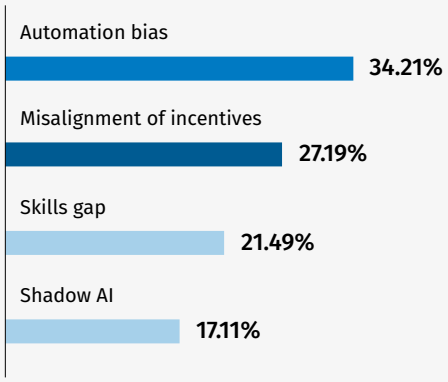
Human automation bias is a governance design problem

Automation Bias leads at 34.21%, with Misalignment of Incentives now second at 27.19%. Automation Bias — the tendency for human reviewers to defer to algorithmic outputs rather than exercise independent judgment — is a failure mode that training alone cannot resolve. It requires governance design in which human review steps carry enough friction, accountability, and decision authority to function as genuine checks, not procedural checkboxes. The combined 61.40 percent for Automation Bias and Misalignment of Incentives confirms that the most consequential AI governance failures are behavioral and cultural, not technical. Skills Gap (21.49%) and Shadow AI (17.11%) trail — not because they are unimportant, but because practitioners are increasingly locating AI risk in the cognitive and organizational dynamics of how AI is used, not merely in the tools themselves.

Advisory signal: Performance management frameworks that reward decision velocity, volume, or throughput without explicit weighting of decision quality and governance compliance are systematically building Automation Bias into AI-augmented workflows. Incentive redesign is a governance intervention — not an HR one.

***Automation bias** has overtaken **misalignment of incentives** as the top human-centric AI risk — a shift that demands governance design changes, not just training programs.*

Survey question: Which human-centric factor poses the greatest risk to your AI safety framework?



Percentage of respondents

61.40%
of concern in behavioral and cultural risk factors — not technical skills gaps or unauthorized tool use.

Source: Trusted AI Survey, April 2026 | 230 respondents

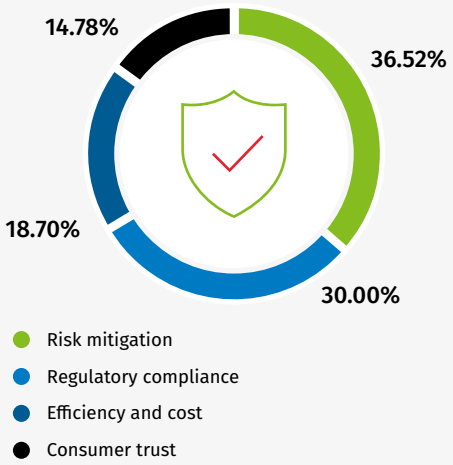
Governing from the inside out

The gap between **risk mitigation** and **regulatory compliance** remains consistent at 6.52 percentage points, reinforcing a durable sector-level pattern: the commitment to Safe AI is being driven more by internal risk appetite than external mandate. This is constructive — it suggests genuine institutional ownership of the governance agenda. However, it also creates a dependency on the accuracy of internal risk perception, which is itself subject to the **automation bias** that practitioners now rank as their top human-centric risk. When cognitive bias toward AI output suppresses internal risk signals, self-motivated governance is a less reliable safeguard than compliance mandates that operate independently of institutional judgment.

Advisory signal: Institutions whose Safe AI programs are primarily internally motivated should ensure that risk assessment processes are insulated from the behavioral pressures that could systematically understate risk. Independent model risk management and internal audit functions play a critical role in maintaining the integrity of that signal — particularly as **automation bias** emerges as the leading human-side risk.

Risk mitigation *— not* ***regulatory compliance*** *— is the primary driver of Safe AI adoption, with a 6.52-point margin suggesting institutions are internalizing AI risk as a genuine institutional concern rather than an externally imposed compliance obligation.*

Survey question: What is the primary driver of 'Safe AI' adoption at your bank?



Risk mitigation outpaces regulatory compliance by 6.52 pts — the sector is governing AI from the inside out.

Source: Trusted AI Survey, April 2026 | 230 respondents

Looking forward: Key implications for US banking institutions

For more information, visit **www.wolterskluwer.com**. Follow us on **LinkedIn**, **Facebook**, **YouTube**, and **Instagram**.

The H1 2026 US Banking AI Risk and Governance — reflecting 230 respondents across institution sizes and seniority levels — confirms a sector at a decisive inflection point. AI adoption in US banking is no longer nascent. It is operational across credit, fraud, compliance, and customer management functions. What remains underdeveloped is the governance infrastructure, organizational readiness, and consumer protection architecture required to sustain that deployment safely as regulators, customers, and counterparties raise their expectations simultaneously.

Fairness and governance controls are co-equal investment priorities. In a meaningful shift, **enhanced fairness** and **bias controls** has emerged as the most urgently needed control investment — edging out **continuous monitoring** in a tightly distributed result. Institutions that have organized their AI control investment around monitoring and governance without concurrent fairness infrastructure are misaligned with both practitioner priority and the direction of regulatory examination focus.

Collections is the immediate consumer protection and fair lending priority. The concentration of perceived consumer harm risk in **collections and recovery** — 35.22 percent, more than 10 points above **credit risk and underwriting** — combined with the agentic AI deployment footprint in that same function creates a specific, near-term exposure profile that demands immediate governance review. Human-in-the-loop design, consistent treatment testing, and customer remediation readiness are non-negotiable prerequisites, not iterative enhancements.

Automation bias is the leading human-side risk — and it requires governance design, not just training. The shift in this wave from **misalignment of incentives** to **automation bias** at the top of the human-centric risk register is a signal that practitioners are locating the greatest AI governance threat not in rogue deployments or skills gaps, but in the cognitive dynamics of how people interact with AI under operating conditions. Governance architecture must account for this — through friction, accountability, and performance frameworks that reward decision quality alongside decision speed.

Banks that align model governance discipline, fairness control investment, consumer protection design, and operational incident readiness are not simply managing risk — they are building the institutional infrastructure that will determine who scales AI safely and who is forced to unwind it.

Wolters Kluwer

wolterskluwer.com